UCLA

UCLA Previously Published Works

Title

Instrumental Variables and Causal Mechanisms: Unpacking the Effect of Trade on Workers and Voters

Permalink

<https://escholarship.org/uc/item/4kp5c6jj>

Authors

Dippel, Christian
Gold, Robert
Heblich, Stephan
et al.

Publication Date

2018-01-17

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-ShareAlike License, available at <https://creativecommons.org/licenses/by-nc-sa/4.0/>

Peer reviewed

Instrumental Variables and Causal Mechanisms: Unpacking the Effect of Trade on Workers and Voters*

Christian Dippel[†] Robert Gold[‡] Stephan Heblich[§] Rodrigo Pinto[¶]

July 28, 2018

Abstract

Instrumental variables (IV) are commonly used to identify treatment effects, but standard IV estimation cannot unpack the complex treatment effects that arise when a treatment and its outcome together cause a second outcome of interest. For example, IV estimations have been used to show that import exposure to low-wage countries has adversely affected high-wage countries' labor markets. They have also been used to show that such import exposure has polarized voters. However, they cannot answer the question to what extent the latter is a consequence of the former. We propose a new identification strategy that allows us to do so, appealing to one additional identifying assumption and requiring no additional instruments. Applying this framework, we estimate that across specifications labor market adjustments explain virtually all of the effect of import exposure on voting. This enables us to provide rigorous evidence that the correct policy response to voter polarization has to be focused on labor markets.

Keywords: Instrumental Variables, Causal Mediation Analysis, Import Exposure, Voting, Local Labor Markets

JEL Codes: C36, D72, F6, J2

*We thank Alberto Alesina, David Autor, Sascha Becker, Sonia Bhalotra, Gilles Duranton, Jon Eguia, Johanna Fajardo, Andreas Ferrara, Markus Frölich, Paola Giuliano, James Heckman, Martin Huber, Kosuke Imai, Ed Leamer, Yi Lu, Craig McIntosh, Ralph Ossa, Sebastian Ottinger, Anne Otto, David Pacini, Bruno Pellegrino, Maria Petrova, Justin Pierce, David Slichter, Daniel Sturm, Peter Schott, Zhigang Tao, Dustin Tingley, Dan Treffer, Nico Voigtländer, Wouter Vermeulen, Till von Wachter, Romain Wacziarg, Frank Windmeijer, Yanos Zylberberg, and seminar participants at Bristol, Kiel, LMU, LSE, Toronto, UCLA, UCSD, U HK, Warwick, the 2015 Quebec Political Economy conference, the 2016 EEA conference, the 2017 VfS conference, and the 2017 INET conference for valuable comments and discussions. We also thank David Slichter for thoughtful comments and Wolfgang Dauth for sharing the crosswalk from product classifications to industry classifications in the German IAB data. Dippel acknowledges financial support from UCLA's Center for Global Management.

[†]University of California, Los Angeles, CCPR, and NBER.

[‡]IfW - Kiel Institute for the World Economy and CESifo.

[§]University of Bristol, CESifo, IZA, and SERC.

[¶]University of California, Los Angeles, CCPR, and NBER.

1 Introduction

Instrumental variables (IV) are broadly used to identify the causal effect of a treatment variable on an outcome in observational data. Standard IV estimation, however, is unable to unpack the causal chain that arises when the treatment and its outcome jointly cause a second outcome of interest. Our empirical application at the nexus of import competition, labor markets, and voting is a case in point: international trade between high- and low-wage countries has risen dramatically in the last thirty years ([Krugman, 2008](#)). Consumers in high-wage countries have benefited from such import exposure through cheaper manufactured goods. However, IV methods have been used to show that import exposure has also caused real wage stagnation and substantial job losses in Western manufacturing ([Autor, Dorn and Hanson, 2013](#); [Dauth, Findeisen and Suedekum, 2014](#); [Malgouyres, 2017](#)).¹ The same IV methods have been used to show that import exposure has increased the support for parties and politicians with protectionist, populist, and nationalist agendas ([Malgouyres, 2014](#); [Dippel, Gold and Heblich, 2015](#); [Feigenbaum and Hall, 2015](#); [Autor, Dorn, Hanson and Majlesi, 2016](#)). While these two findings seem to be related, standard IV methods cannot quantify to what extent import exposure has turned voters toward political populism because such exposure adversely affected labor markets. Unpacking this causal mechanism is important to guide policy-makers in designing effective responses to political populism.

From an econometric perspective, we investigate the problem of identifying causal relations when an endogenous treatment and its outcome together cause a second outcome of interest. We propose a solution to the problem that does not require additional instrumental variables and can be easily implemented using the well-known two-stage least squares (2SLS) estimator. We begin by clarifying the identification challenge. The starting point is to estimate the effect of a non-random treatment T (e.g. import exposure) on an outcome M (e.g. regional labor market adjustments). The ordinary least squares (OLS) estimate of said treatment effect may be biased by omitted variables that affect both T and M (e.g. regional demand shocks may reduce regional imports as well as employment). The solution involves using an instrumental variable Z that affects T (i.e. there is a first-stage relation) but is uncorrelated with the omitted variables (i.e. the

¹Not every paper that addresses this question uses IV. See for example [Dix-Carneiro and Kovak \(2017\)](#) and [Pierce and Schott \(2016\)](#), who instead focus on exogenous variation in tariff reductions.

Table 1: The Identification Problem of Mediation Analysis with IV

A. Graphical Representation

B. Model Equations

$T = f_T(Z, \epsilon_T)$	$T = f_T(Z, \epsilon_T)$	$T = f_T(Z, \epsilon_T), M = f_M(T, \epsilon_M)$
$M = f_M(T, \epsilon_M)$	$Y = g_Y(T, \eta_Y)$	$Y = f_Y(T, M, \epsilon_Y)$
$Z \perp\!\!\!\perp (\epsilon_T, \epsilon_M)$	$Z \perp\!\!\!\perp (\epsilon_T, \eta_Y)$	$Z \perp\!\!\!\perp (\epsilon_T, \epsilon_M, \epsilon_Y)$

Notes: (a) Model I is the standard IV model, which enables the identification of the causal effect of T on M . Model II is the standard IV model that enables the identification of the causal effects of T on Y . Model III is the IV Mediation Model with an instrumental variable Z . (b) Panel A gives the graphical representation of the models. Panel B presents the non-parametric structural equations of each model. Conditioning variables are suppressed for sake of notational simplicity. We use $\perp\!\!\!\perp$ to denote statistical independence.

exclusion restriction holds).² This is the standard IV solution and is depicted in *Model I* in Table 1. T is endogenous in a regression of M on T (i.e. $\epsilon_T \not\perp\!\!\!\perp \epsilon_M$), but Z is exogenous (i.e. $Z \perp\!\!\!\perp \epsilon_T, \epsilon_M$).

We are interested in the identification challenge that arises when there is a second outcome of interest Y (e.g. voting) that is likely to be caused by T , both through M as well as “directly” (which is to say, through other channels). The most straightforward approach to this is to simply estimate the ‘total effect’ of T on Y using the same IV approach, as depicted in *Model II* in Table 1: $\epsilon_T \not\perp\!\!\!\perp \eta_Y$, but Z is exogenous (i.e. $Z \perp\!\!\!\perp \epsilon_T, \eta_Y$).³ In combination, *Model I* and *Model II* estimate the causal effect of T on M and the causal effect of T on Y . However, this does not identify to what extent the former causes the latter. In our empirical setting, there clearly could be other channels that directly link T to Y : On the one hand, if import exposure created anxiety about the future, this may by itself turn voters toward populism (Mughan and Lacy, 2002; Mughan,

²We base our analysis on the IV strategy proposed by Autor et al. (2013), which instruments for one high-wage country’s trade exposure to China with other high-wage countries’ industry-specific trade exposure to China.

³It is common to use the same instrument to identify the causal effect of a treatment on several outcomes, and the application studied here is no different. For example, three pairs of papers in the related literature each use the same identification strategy to separately investigate the effect of import exposure on labor markets and on some form of political outcomes; e.g. Autor et al. (2013) and Autor et al. (2016), Malgouyres (2017) and Malgouyres (2014), as well as Pierce and Schott (2016) and Che, Lu, Pierce, Schott and Tao (2016).

Bean and McAllister, 2003).⁴ On the other hand, import exposure may be politically moderating if it lowered consumer prices, if targeted government transfers like *Trade Adjustment Assistance* increased, if manufacturers shifted production toward more differentiated higher-markup output varieties (as in Holmes and Stevens, 2014), or if it lead to task-upgrading within industries and occupations (as in Becker and Muendler, 2015). Depending on these factors' relative importance, their aggregate effect on support for populists may be positive or negative. If these direct effects as a whole were negative (i.e. the direct effect has the opposite sign of the total effect) the effect of import exposure on voting that is mediated by labor market adjustments would actually be larger than the total effect estimated by *Model II*. We find evidence that this is indeed the case.

The identification challenges that arise from this discussion are depicted in *Model III* in Table 1. Equations $M = f_M(T, \epsilon_M)$ and $Y = f_Y(T, M, \epsilon_Y)$ imply that T causes Y indirectly through M as well as directly, i.e. through other channels (that are graphically represented by the arrow directly linking T to Y). In a regression of Y on both T and M , there are two endogenous regressors (i.e. $\epsilon_T \not\perp \epsilon_Y$, $\epsilon_M \not\perp \epsilon_Y$), but there is only one instrument Z to address this endogeneity. *Model III* is a *mediation model*, i.e. one where T (import exposure) causes an intermediate outcome M (labor market adjustments) that is also a *mediator* in T 's effect on a final outcome Y (voting).⁵ Most of the approaches to identification in mediation analysis assume that T is as good as randomly assigned (i.e. $\epsilon_T \perp \epsilon_M$), making them not applicable to the IV settings we are interested in. See, e.g., Imai, Keele, Tingley and Yamamoto (2011). The only existing approaches to achieving identification in the IV setting of *Model III* require separate dedicated instruments for M , which require additional exogeneity assumptions that are considerably more restrictive than the standard ones (e.g. Jun, Pinkse, Xu and Yildiz 2016; Frölich and Huber 2017).

Our proposed solution does not assume away endogeneity in any of the key relationships in *Model III* and does not require additional instruments. Instead, we rely on the insight that in many research settings the omitted variable concerns themselves suggest a natural solution. This is the case when T is endogenous in a regression of Y on T primarily because of omitted variables that affect M . For example, in the literature above, the main endogeneity concern in a regression

⁴ In our empirical specification, import exposure refers to the net import exposure, i.e. the exposure to imports *after* accounting for export exposure.

⁵ Mediation analysis decomposes the *total effect* of T on Y into the *indirect effect* of T on Y that operates through M and the *direct effect* that does not. The indirect effect may alternatively be labeled as the '*mediated effect*'. For recent works on this literature, see Heckman and Pinto (2015b); Pearl (2014); Imai, Keele and Tingley (2010a).

of regional manufacturing employment (M) on import exposure (T) is that unobserved adverse regional demand shocks reduce regional imports as well as employment, and it seems likely that such shocks affect voting (Y) primarily to the extent that they affect labor markets. We show that this assumption alone is sufficient to unpack the causal channels in *Model III*, allowing us to identify the extent to which T causes Y through M . We further show that under linearity, the resulting identification framework is straightforwardly estimated using three separate 2SLS estimations of the effect of T on M , the effect of T on Y , and the effect of M on Y conditional on T .⁶ We also develop a procedure to bound the possible range of the direct and the indirect effects linking T and Y when the identifying assumption of our framework is relaxed.

We apply our method to data on regional import exposure, labor market adjustments, and voting patterns in Germany from 1987 to 2009. The data is organized as a stacked panel of two first differences for the periods 1987–1998 and 1998–2009, with specific start- and end-points dictated by national election dates. The analysis precedes the European debt crisis and each period includes a large international trade shock: In 1989, the fall of the Iron Curtain opened up the Eastern European markets, and in 2001 China’s accession to the WTO led to another large increase in import exposure. We use German data in part because it offers several advantages, especially relative to the U.S., for the question at hand: (i) Germany’s multiparty system straddles the entire political spectrum, from the far left to the extreme right, so that we can consistently measure changes in political preferences over time. (ii) Germans cast their main vote for a party at large as opposed to by district, so local voting patterns are not confounded by local variation in political messaging.⁷ (iii) We are able to measure vote shares, regional import exposure, and labor market conditions all at the same statistical unit of 408 *Landkreise*.⁸ (iv) Unique among attitudinal socioeconomic surveys, the German *Socio-Economic Panel*’s (SOEP) long-running panel structure allows us to cross-validate the aggregate results with an individual-level panel-analysis, relating decadal

⁶ Our main focus lies on estimating the product of the effect T on M and the effect M on Y , i.e. the indirect effect, and to compare it to the estimated total effect of T on Y . This objective is naturally similar to traditional approaches to mediation analysis, which assume that both T and M are exogenous, and apply OLS to estimate three equations,

$$Y = \delta_Y^T \cdot T + \eta_Y, \quad M = \beta_M^T \cdot T + \epsilon_M, \quad \text{and} \quad Y = \beta_Y^T \cdot T + \beta_Y^M \cdot M + \epsilon_Y,$$

and then compare the total effect δ_Y^T to the indirect effect $\beta_Y^M \cdot \beta_M^T$. See [Baron and Kenny \(1986\)](#) and [MacKinnon \(2008\)](#) for an overview, as well as the discussion in our section 2.1.

⁷ This vote, called *Zweitstimme*, mainly determines a party’s share of parliamentary seats.

⁸ In U.S. data, by contrast, one observes vote-shares in 3,007 counties, politicians in 435 congressional districts, and trade shocks in 741 commuting zones.

changes in individual workers' stated party preferences to changes in their local labor market's import exposure over the same time.

We combine changes in national sector-specific trade flows with regional labor markets' initial industry mix to determine regional import exposure (T), and instrument T with a measure based on other high-wage countries' sector-specific trade flows (Z). Estimating *Model I* in Table 1, we corroborate existing results that import exposure significantly reduces total employment (M), particularly in manufacturing, raises unemployment, and negatively affects manufacturing wages. Estimating *Model II* in Table 1, we find that import exposure (T) increased voter polarization (Y). There is a significant positive effect on the vote share of the nationalist and highly protectionist extreme right.⁹ There are no significant effects on turnout, or on any of the mainstream parties, the far left, or other small parties.¹⁰ These findings are corroborated by the SOEP's individual-level data, where we can show that the effects are entirely driven by low-skill workers employed in manufacturing, i.e. those most affected by the labor market adjustments to increasing import exposure. Using gravity residuals instead of our IV strategy yields similar results.¹¹

Next, we estimate *Model III*. The effect of trade-induced labor market adjustments on voting is at least as large as the total effect of import exposure on voting, suggesting that other channels through which import exposure affects voting are moderating the total effect. When we relax the identifying assumptions and derive bounds on the indirect effect, we find a lower bound of 70 percent. This, even at the lower bound, labor market adjustments appear to be the primary reason for the populist backlash against import exposure.

Our paper's contribution is twofold: It addresses a relevant substantive question at the nexus of the literatures on trade, local labor markets, and politics; in order to do so, it makes a methodological contribution to the literature on causal mechanisms and on IV. On the methodological side, we offer a mediation model that relies on a single instrumental variable Z that directly causes T to identify three causal effects, while allowing for endogenous variables caused by confounders

⁹ Election outcomes are divided into changes in the vote-share of (i) four mainstream parties: the CDU, the SPD, the FDP and the Green party, (ii) extreme-right parties, (iii) far-left parties, (iv) other small parties, and (v) turnout, see Falck, Gold and Heblich (2014).

¹⁰ Of course, *all other parties combined* must have a vote share loss of equal magnitude and significance to the extreme right's gain.

¹¹ We report these for completeness, as this is standard in the literature on import exposure and labor markets (Autor et al., 2013; Dauth et al., 2014). However, our focus is naturally on the IV setting to which our identification framework applies.

and for unobserved mediators. This parsimonious feature is useful for the typical observational data setting, where good instrumental variables are scarce. Our model can be estimated by well-known 2SLS methods, its identifying assumption can be relaxed to derive bounds instead of point estimates, and it can be applied to a potentially broad range of empirical research questions in which an endogenous treatment and its primary outcome together cause a second outcome of interest. On the substantive side, our analysis confirms that labor market adjustments, concentrated in manufacturing, are the main reason for the political backlash against free trade in the data we study.

The rest of the paper proceeds as follows: Section 2 explains our identification approach. Section 3 describes the data. Section 4 presents the IV results for *Model I* and *Model II*, establishing the causal effects of import exposure on labor markets and voting. Section 5 applies *Model III*. Section 6 concludes.

2 Mediation with an Instrumental Variable

In the following, section 2.1 lays out the definition and identification of causal effects in a mediation model, i.e. a model where a treatment T and its outcome M jointly cause another outcome Y . In section 2.2, we present the exclusion restrictions needed to obtain identification of all causal effects using an instrument Z for T . These include the standard IV assumptions needed to identify *Model I* and *Model II*, as well as the novel exclusion restriction needed to unpack *Model III*. Under the assumption of linearity, section 2.3 derives an easily operationalized estimation procedure that relies on a series of 2SLS estimations. Finally, in section 2.4 we derive a complementary bounding exercise that relaxes our identifying assumption.

2.1 Causal Effects in the Mediation Model

Our goal is to evaluate a sequence of causal relations where import exposure T causes labor market adjustments M , and both T and M cause changes in voting behavior Y . Such a sequence of causal relations is called a mediation model (Pearl, 2011). In a mediation model, the total effect (TE) of T on Y can be decomposed as the sum of the ‘indirect effect’ (IE) of T on Y that is mediated by M and the ‘direct effect’ (DE) of T on Y that is not mediated by M . Let $\epsilon_T, \epsilon_M, \epsilon_Y$ be the unobserved

error terms associated with variables T, M, Y , respectively.

A general nonparametric model that portrays these causal relations is then given by

$$T = f_T(\epsilon_T), \quad (1)$$

$$M = f_M(T, \epsilon_M), \quad (2)$$

$$Y = f_Y(T, M, \epsilon_Y). \quad (3)$$

Causal effects are defined by the difference between counterfactual variables.¹² For instance, let $M(t)$ be the counterfactual variable M when T takes the value t . The causal relation in equation (2) yields the counterfactual variable $M(t)$; while causal relation (3) yields the counterfactual variables $Y(t)$, $Y(m)$, and $Y(m, t)$ as defined below:¹³

$$M(t) = f_M(t, \epsilon_M), \quad (4)$$

$$Y(t) = f_Y(t, M(t), \epsilon_Y), \quad (5)$$

$$Y(m) = f_Y(T, m, \epsilon_Y), \quad (6)$$

$$Y(m, t) = f_Y(t, m, \epsilon_Y). \quad (7)$$

The causal effect of T on M when T takes values t_1, t_0 is given by the expected value of the difference $E(M(t_1) - M(t_0))$. The three main causal effects of interest in a mediation setting (i.e. the TE , DE , and IE) are defined as follows.¹⁴

$$TE = E(Y(t_1) - Y(t_0)) \equiv \int E(Y(t_1, M(t_1)) - Y(t_0, M(t_0))) \quad (8)$$

$$DE(t) = E(Y(t_1, M(t)) - Y(t_0, M(t))) \equiv \int E(Y(t_1, m) - Y(t_0, m)) dF_{M(t)}(m) \quad (9)$$

$$IE(t) = E(Y(t, M(t_1)) - Y(t, M(t_0))) \equiv \int E(Y(t, m)) [dF_{M(t_1)}(m) - dF_{M(t_0)}(m)] \quad (10)$$

The seminal work of [Robins and Greenland \(1992\)](#) showed that under the assumption of the mutual independence of the error terms, i.e. $\epsilon_T \perp\!\!\!\perp \epsilon_M, \epsilon_M \perp\!\!\!\perp \epsilon_Y, \epsilon_T \perp\!\!\!\perp \epsilon_Y$, the TE , IE , and DE can be causally identified in equations (1)–(3). See [Online Appendix A](#). Under linearity, this

¹² Causal effects and counterfactual outcomes, as well as exclusion restrictions are defined non-parametrically. As is standard, we therefore use a non-parametric model when discussing causality in sections 2.1 and 2.2. Counterfactual variables are defined by *fixing* an argument of a structural equation to a value. The counterfactual variable $M(t)$, for instance, is defined by *fixing* the T -input of this structural equation to the value t , namely, $M(t) = f_M(t, \epsilon_M)$. See [Heckman and Pinto \(2015a\)](#) for a discussion.

¹³ Counterfactual variable $Y(t)$ in (5) denotes the potential outcome Y when T takes the value t , $Y(m)$ is the counterfactual outcome when M is fixed to the value m , and $Y(m, t)$ is the counterfactual outcome that arises when both T and M are fixed to t and m , respectively.

¹⁴ [Pearl \(2011\)](#) makes a distinction between controlled (or “prescriptive”) and natural (or “descriptive”) effects. He uses the terms *controlled* direct effect and *controlled* indirect effect for equations (9) and (10) respectively. In the linear case, which will be our focus, this distinction disappears. See also footnote 21.

amounts to the series of three OLS regressions in footnote 6.¹⁵

The assumption of the mutual independence among all error terms is clearly a strong one: Assuming that the error terms ϵ_T and ϵ_M are statistically independent, i.e. $\epsilon_T \perp\!\!\!\perp \epsilon_M$, implies that there are no unobserved variables that jointly cause T and M . In this case, T is exogenous with respect to M , i.e. we have random assignment as in a randomized control trial (RCT), a condition that is rarely satisfied in observational data. For example, in our empirical application, the literature's concern that unobserved adverse regional demand shocks reduce regional imports (T) as well as employment (M) implies $\epsilon_T \not\perp\!\!\!\perp \epsilon_M$, which precludes the independence between T and counterfactuals $M(t)$ and $Y(t)$ in equations (4), (5), i.e. $T \not\perp\!\!\!\perp (Y(t), M(t))$. Hence, T is endogenous and the causal effect of T on the outcomes M and Y cannot be identified on the basis of their observed distributions. This is the identification challenge that *Model I* and *Model II* in Table 1 solve using IV.¹⁶

Relatedly, assuming the mutual independence among all error renders both the *DE* and *IE* identified because it makes variables T and M exogenous with respect to Y (See [Online Appendix A](#)). This assumption does not follow even when treatment T is randomly assigned, as in an RCT. Drawing further on our empirical application to illustrate the pitfalls of this assumption, it is likely that $\epsilon_M \not\perp\!\!\!\perp \epsilon_Y$ because unobserved industry or worker characteristics that affect labor market outcomes may also affect political preferences. The statistical dependence $\epsilon_M \not\perp\!\!\!\perp \epsilon_Y$ precludes the independence between M and counterfactuals $(Y(m, t), Y(m))$ in equations (6), (7), i.e. $M \not\perp\!\!\!\perp (Y(m), Y(m, t))$. Hence, the causal effect of M on Y cannot be identified on the basis of their observed distributions. This is the additional identification challenge that arises when trying to estimate *Model III* in Table 1, which our framework will allow us to solve by using an instrumental variable.

In summary, the mutual independence among all error terms ϵ_T, ϵ_M , and ϵ_Y is unlikely to hold in observational data. However, without this assumption causal effects are not identified on the basis of observed distributions. To gain identification, we modify the standard mediation

¹⁵ A large literature on mediation analysis relies on the Sequential Ignorability Assumption [A-3](#) of [Imai, Keele and Yamamoto \(2010b\)](#) to identify mediation effects. We discuss this assumption in [Online Appendix B](#), which is related to the assumption of the mutual independence of the error terms $\epsilon_T, \epsilon_M, \epsilon_Y$. See [Frölich and Huber \(2017\)](#) for a recent review of the mediation literature.

¹⁶ *Model II* in Table 1 amounts to a direct estimate of the *TE* in equation (8). By contrast, *Model III* in Table 1, which makes the mediation setup explicit, estimates the *TE* as the sum of the *DE* and *IE* in equations (9)–(10).

model by adding an instrument Z that causes T , thus equation (1) becomes $T = f_T(Z, \epsilon_T)$. Stated more succinctly, we study identification of the mediation model in an IV setting. In section 2.2 we derive two well-known exclusion restrictions and one novel one, which together allow for the identification of all causal effects of interest while allowing for the dependence among all key error terms ϵ_T, ϵ_M , and ϵ_Y , and requiring an instrument only for T .

2.2 Exclusion Restrictions

In the following, section 2.2.1 discusses the well-known exclusion restrictions needed for identification of *Model I* and *Model II*; section 2.2.2 presents the novel exclusion restriction needed for the identification of *Model III*, without the use of additional instruments.

2.2.1 Exclusion Restriction to Identify *Model I* and *Model II* in Table 1

The standard property of an instrumental variable Z is that it affects outcomes M and Y only through its impact on treatment T . Put another way, for Z to be an instrument, it must be the case that Z is statistically independent of unobserved error terms ϵ_T, ϵ_M and ϵ_Y , which jointly cause T, M and Y . This property is of course well-established, but in the interest of clarity we state it again as Assumption A-1.¹⁷

Assumption A-1 *The independence relation $Z \perp\!\!\!\perp (\epsilon_T, \epsilon_M, \epsilon_Y)$ holds in the mediation model (1)–(3).*

Assumption A-1 merely states the independence condition that characterizes Z as an instrumental variable for T in $M(t)$ as well as $Y(t)$. Lemma L-1 states the well-established result that Assumption A-1 generates the two exclusion restrictions needed to estimate *Model I* and *Model II*.¹⁸

Lemma L-1 *Under Assumption A-1, the following statistical relations hold:*

Targeted Causal Relation	IV Relevance (First Stage)		Exclusion Restriction
$T \rightarrow Y$	$Z \not\perp\!\!\!\perp T$	and	$Z \perp\!\!\!\perp Y(t)$
$T \rightarrow M$	$Z \not\perp\!\!\!\perp T$	and	$Z \perp\!\!\!\perp M(t)$

Proof P-1 *See P-1 in Appendix A.*

¹⁷ As mentioned, Autor et al. (2013) fostered a large literature that uses import exposure of other countries (say O 's) as an instrument Z for one specific country's (say G 's) import exposure to low-wage manufacturing countries. See expressions (35) and (36) for the precise definitions of T and Z in our empirical application.

¹⁸ *Model I* stands for the standard IV model that evaluates the effect of T on M using Z as the instrument. *Model II* stands for the IV model that evaluates the total effect (TE) of T on Y .

The exclusion restrictions in **L-1** imply that Z is a valid instrument for T , and thus the counterfactual outcomes $M(t)$ and $Y(t)$ can be evaluated using standard IV techniques. This is not surprising. **L-1** simply means that an instrument for T enables the identification of the causal effect of T on M as well as T on Y . The exclusion restriction that applies to the mediator M also applies to outcome Y . The fact that M causes Y (and not the opposite) plays no role in generating the exclusion restrictions. Indeed, the lemma would remain the same if the causal relation $M \rightarrow Y$ were reversed to $M \leftarrow Y$. This irrelevance of the causal direction between M and Y highlights the fact that identification of *Model I* and *Model II* does not lead to identification of *Model III*. In Section 2.2.2 we address this problem by invoking the error dependence structure discussed therein.

Remark 2.1 Exclusion restrictions such as **L-1** are necessary but not sufficient to identify causal effects. An extensive IV literature exists on the additional assumptions that grant the identification of causal effects.¹⁹

2.2.2 Exclusion Restriction to Identify *Model III* in Table 1

Our interest is in settings where T and M jointly cause a second outcome of interest Y , as in equation (3). What we will exploit for identification is that in many such settings the omitted variable concerns themselves suggest a natural solution to the identification challenges involved. This is best illustrated in our empirical application, where the main endogeneity concern in a regression of regional manufacturing employment (M) on import exposure (T) is that unobserved adverse regional demand shocks reduce regional imports as well as employment. The critical observation is that the endogeneity concerns in the relation between import exposure and voting (Y) emanate primarily from the labor market channel. In other words, the main concern about unobservables in this relation is that regional demand shocks that affect import exposure also affect voting through their labor market effects. Using this intuition, the framework we propose hinges on allowing unobserved shocks that impact import exposure to also affect voter preferences through labor markets, but not through **unrelated** channels. This is made precise in Assumption **A-2**.

¹⁹ See [Dahl, Huber and Mellace \(2017\)](#) for a recent review. Examples of these additional assumptions are monotonicity ([Imbens and Angrist, 1994](#); [Heckman and Pinto, 2017](#)), separability of the choice equation ([Heckman and Vytalil, 2005](#)) or control functions ([Blundell and Powell, 2003, 2004](#)).

Assumption A-2 The following dependence relations hold in the mediation model (1)–(3):

$$\epsilon_T \perp\!\!\!\perp \epsilon_Y, \text{ but } \epsilon_T \not\perp\!\!\!\perp \epsilon_M, \epsilon_M \not\perp\!\!\!\perp \epsilon_Y, \text{ and } \epsilon_T \not\perp\!\!\!\perp \epsilon_Y|\epsilon_M \quad (11)$$

The only assumption in A-2 that is needed for identification is that error terms ϵ_T, ϵ_Y are unconditionally independent, i.e. $\epsilon_T \perp\!\!\!\perp \epsilon_Y$. None of the remaining assumptions $\epsilon_T \not\perp\!\!\!\perp \epsilon_M$, $\epsilon_M \not\perp\!\!\!\perp \epsilon_Y$, and $\epsilon_T \not\perp\!\!\!\perp \epsilon_Y|\epsilon_M$ is needed for identification. These remaining assumption merely clarify that we do not assume away endogeneity in any of the key relations: We allow for error terms ϵ_T, ϵ_M to correlate, we allow for error terms ϵ_M, ϵ_Y to correlate, and we allow for ϵ_T, ϵ_Y to correlate conditional on ϵ_M , i.e. $\epsilon_T \not\perp\!\!\!\perp \epsilon_Y|\epsilon_M$, which implies M is endogenous in equation (3).

Remark 2.2 Under $\epsilon_T \perp\!\!\!\perp \epsilon_Y$, either $\epsilon_T \perp\!\!\!\perp \epsilon_M$ or $\epsilon_M \perp\!\!\!\perp \epsilon_Y$ would imply $\epsilon_T \perp\!\!\!\perp \epsilon_Y|\epsilon_M$:

- (i) If $\epsilon_T \perp\!\!\!\perp \epsilon_M$ and $\epsilon_M \not\perp\!\!\!\perp \epsilon_Y$, then the causal effects of T on M and T on Y (Model I and Model II) could be identified using OLS. However, identification of the causal effect of M on Y would require a separate dedicated instrument $\tilde{Z}$ that has no impact on Y other than through M .
- (ii) If $\epsilon_T \not\perp\!\!\!\perp \epsilon_M$ and $\epsilon_M \perp\!\!\!\perp \epsilon_Y$, then the causal effects of T on M and T on Y could only be identified using IV. However, the causal effect of M on Y could be identified using OLS.
- (iii) Finally, if $\epsilon_T \perp\!\!\!\perp \epsilon_M$ and $\epsilon_M \perp\!\!\!\perp \epsilon_Y$, then the causal relations (1)–(3) could be identified by the three OLS regressions in footnote 6.

It is important to note that we are not claiming that the assumption $\epsilon_T \perp\!\!\!\perp \epsilon_Y$ is appropriate in all mediation-type settings characterized by causal relations (1)–(3). Instead, our identification approach is appropriate for settings like ours where T is endogenous in the relation of T and Y primarily because of omitted variables that affect M , and through M also Y . There will be settings where this assumption is implausible or at least debatable. It is therefore desirable to be able to relax $\epsilon_T \perp\!\!\!\perp \epsilon_Y$. We do this in Section 2.4, where we show that under $\epsilon_T \not\perp\!\!\!\perp \epsilon_Y$ we can still bound the correlations between error terms and the estimates of all key causal relations in the data.

A-2 yields a non-obvious new exclusion restriction, which allows for the use of Z as an instrument for M , when conditioned on T , as stated in Lemma L-2. Critically, this implies that we can identify three causal effects using only a single instrument (or set of instruments) dedicated for T .

Lemma L-2 Under A-1–A-2, the following statistical relation holds:

Targeted Causal Relation	IV Relevance (First Stage)	Exclusion Restriction
for $M \rightarrow Y$	$Z \not\perp\!\!\!\perp M T$	and $Z \perp\!\!\!\perp Y(m) T$

Proof P-2 See P-2 in Appendix A.

Lemma **L-2** is novel, and it is worth discussing both Z 's explanatory power for M , conditional on T , as well as Z 's validity as an instrument for M , conditional on T : The relevance of the IV for M conditional on T , i.e. $Z \not\perp\!\!\!\perp M|T$, comes from a residual variation argument, which we again illustrate in our empirical application: there, the endogeneity concern in the relation between T and M is that an industry- j -specific domestic demand shock will reduce both local import exposure (T) and local employment (M) in regions that are specialized in industry j . The solution advanced in the literature is to use other (high-wage) countries' imports as the basis of an instrument (Z) that is orthogonal to Germany-specific demand conditions. The key piece of intuition for Z 's explanatory power for M , conditional on T , is that when other countries' industry-specific imports are high relative to (i.e. conditional on) German industry-specific imports, these will partly reflect the unobserved German industry-specific demand conditions, i.e. the source of the bias. Conditional on German imports (T), other countries' higher imports in a given sector therefore "cause" additional reductions in German employment because they reflect negative German demand conditions.²⁰ The exclusion restriction of **L-2** implies that the instrumental variable Z can be used to evaluate the causal relation of M on Y if (and only if) conditioned on T . Indeed, while $Z \perp\!\!\!\perp Y(m)|T$ holds, $Z \perp\!\!\!\perp Y(m)$ does not. In other words, Z is a valid instrument for M only when conditioning on T . In **Online Appendix C** we summarize the error dependences of **A-2** once more visually in a directed acyclic graph (DAG), as in *Model III* in Table 1.

Remark 2.3 The exclusion restrictions in **L-1** and **L-2** also hold for a more general model that allows for an unobserved mediator U that is caused by T and causes both M and Y . Notationally, this model is characterized by the following equations: $T = f_T(Z, \epsilon_T)$, $U = f_U(T, \epsilon_U)$, $M = f_M(T, U, \epsilon_M)$, and $Y = f_Y(T, M, U, \epsilon_Y)$. We investigate this model in **Online Appendix D**.

Corollary C-1 Under **A-1–A-2**, the counterfactual outcome $Y(m)$ conditioned on $T = t$ is equal in distribution to the counterfactual outcome $Y(m, t)$, i.e., $(Y(m)|T = t) \stackrel{d}{=} Y(m, t)$.

Proof P-3 See **P-3** in **Appendix A**.

Corollary **C-1** states that the counterfactual outcome $Y(m)$ conditioned on $T = t$, i.e. $(Y(m)|T = t)$, is equal in distribution to the counterfactual outcome $Y(m, t)$. As a consequence, the total effect (equation (8)) can be decomposed into the direct effect (equation (9)), and the indirect effect (equation (10)).

²⁰ Whether they indeed do so is a question of explanatory power, not identification.

Remark 2.4 An alternative identification approach to identify the counterfactual outcome $Y(m)$ is with an additional dedicated IV. Consider an additional variable $\tilde{Z}$ that plays the role of an IV that is exclusively dedicated to M . This means that variable $\tilde{Z}$ is characterized by two properties: (i) $\tilde{Z}$ does not cause T ; and (ii) $\tilde{Z}$ has no impact on Y other than through M . This instrument could be used to evaluate the causal effect of M on Y ; however, the availability of such instrument is unlikely in most empirical settings. See e.g. [Frölich and Huber \(2017\)](#), and also [Online Appendix E](#), where we discuss the potential applicability as well as pitfalls of using a second dedicated instrument $\tilde{Z}$.

2.3 The IV Mediation Model under Linearity

As is standard, in section 2.2 we have used a non-parametric model to discuss causality. To derive an easily operationalized estimation procedure from the identification results we derived, we will assume linearity from here on. Linearity is commonly assumed in many applied literatures, in particular is it pervasive in the local labor markets literature which our empirical application belongs to. Under linearity, the exclusion restrictions in section 2.2 simply translate into a lack of correlation between error terms. We will show that this renders a just-identified model. We will show that we can use standard 2SLS estimation procedures to identify the causal effect of T on M , T on Y , and M on Y , thereby decomposing the total effect of T on Y into its direct and indirect effect. Under linearity (and adding the instrument), the causal relations in equations (1)–(3) can be written as follows:

$$Z = \epsilon_Z, \quad (12)$$

$$T = \beta_T^Z \cdot Z + \epsilon_T, \quad (13)$$

$$M = \beta_M^T \cdot T + \epsilon_M, \quad (14)$$

$$Y = \beta_Y^T \cdot T + \beta_Y^M \cdot M + \epsilon_Y, \quad (15)$$

where $\epsilon_Z, \epsilon_T, \epsilon_M, \epsilon_Y$, are error terms whose variances are denoted by $\sigma_{\epsilon_Z}^2, \sigma_{\epsilon_T}^2, \sigma_{\epsilon_M}^2, \sigma_{\epsilon_Y}^2$, respectively. Let ρ_{TM} stand for the correlation between ϵ_T and ϵ_M . Likewise, let ρ_{TY}, ρ_{MY} stand for the correlations between ϵ_T, ϵ_Y and ϵ_M, ϵ_Y , respectively. For sake of notational simplicity, we assume that each variable has a mean of zero. This assumption does not incur a loss of generality. The direct effect is given by the coefficient $DE = \beta_Y^T$, the indirect effect is given by the coefficient multiplication $IE = \beta_M^T \cdot \beta_Y^M$, and the total effect is the sum of these two terms $TE = \beta_Y^T + \beta_M^T \cdot \beta_Y^M$.²¹

²¹ The linear model lacks treatment-mediator interactions and therefore has homogeneous effects, i.e. [Pearl's](#) nat-

We are interested in evaluating the linear coefficients $\beta_T^Z, \beta_M^T, \beta_Y^T, \beta_Y^M$. The identification of these coefficients depends on the covariance matrix of observed data. Therefore, it is useful to represent equations (12)–(15) in matrix form. Let $\mathbf{X} = [Z, T, M, Y]'$ be the vector of observed random variables and $\epsilon = [\epsilon_Z, \epsilon_T, \epsilon_M, \epsilon_Y]'$ be the vector of unobserved error terms. Matrix Ψ in (16) stands for the arrangement of linear coefficients.

$$\underbrace{\begin{bmatrix} Z \\ T \\ M \\ Y \end{bmatrix}}_{\mathbf{X}} = \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 \\ \beta_T^Z & 0 & 0 & 0 \\ 0 & \beta_M^T & 0 & 0 \\ 0 & \beta_Y^T & \beta_Y^M & 0 \end{bmatrix}}_{\Psi} \cdot \underbrace{\begin{bmatrix} Z \\ T \\ M \\ Y \end{bmatrix}}_{\mathbf{X}} + \underbrace{\begin{bmatrix} \epsilon_Z \\ \epsilon_T \\ \epsilon_M \\ \epsilon_Y \end{bmatrix}}_{\epsilon}. \quad (16)$$

Equations (12)–(15) are thus written as $\mathbf{X} = \Psi \cdot \mathbf{X} + \epsilon$. Equation (17) presents the covariance matrix $\Sigma_{\mathbf{X}}$ of observed variables $\mathbf{X}$:

$$\Sigma_{\mathbf{X}} \equiv \text{Var} \begin{pmatrix} Z \\ T \\ M \\ Y \end{pmatrix} = \begin{bmatrix} \sigma_{ZZ} & \sigma_{ZT} & \sigma_{ZM} & \sigma_{ZY} \\ \cdot & \sigma_{TT} & \sigma_{TM} & \sigma_{TY} \\ \cdot & \cdot & \sigma_{MM} & \sigma_{MY} \\ \cdot & \cdot & \cdot & \sigma_{YY} \end{bmatrix}. \quad (17)$$

Let Σ_{ϵ} denote the covariance matrix of unobserved error terms ϵ . Assumption A-1 states that Z is an IV. It implies that error term ϵ_Z is statistically independent of $\epsilon_T, \epsilon_M, \epsilon_Y$. Thus, Σ_{ϵ} is given by

$$\Sigma_{\epsilon} \equiv \text{Var} \begin{pmatrix} \epsilon_Z \\ \epsilon_T \\ \epsilon_M \\ \epsilon_Y \end{pmatrix} = \begin{bmatrix} \sigma_{\epsilon_Z}^2 & 0 & 0 & 0 \\ \cdot & \sigma_{\epsilon_T}^2 & \rho_{TM}\sigma_{\epsilon_T}\sigma_{\epsilon_M} & \rho_{TY}\sigma_{\epsilon_T}\sigma_{\epsilon_Y} \\ \cdot & \cdot & \sigma_{\epsilon_M}^2 & \rho_{MY}\sigma_{\epsilon_M}\sigma_{\epsilon_Y} \\ \cdot & \cdot & \cdot & \sigma_{\epsilon_Y}^2 \end{bmatrix}. \quad (18)$$

Assumption A-2 states that $\epsilon_T \perp\!\!\!\perp \epsilon_Y$, but allows for $\epsilon_T \not\perp\!\!\!\perp \epsilon_M$, $\epsilon_M \not\perp\!\!\!\perp \epsilon_Y$, and $\epsilon_T \not\perp\!\!\!\perp \epsilon_Y | \epsilon_M$. Under linearity, this translates into $\rho_{TY} = 0$, but $\rho_{TM} \neq 0$ and $\rho_{MY} \neq 0$ in Σ_{ϵ} . As well, $\epsilon_T \not\perp\!\!\!\perp \epsilon_Y | \epsilon_M$ implies $\rho_{TY | \epsilon_M} \neq 0$. In Appendix B, we describe how to straightforwardly generate a simulated dataset with these dependence relations.

The identification of linear coefficients $\beta_T^Z, \beta_M^T, \beta_Y^T, \beta_Y^M$ and the unobserved parameters in Σ_{ϵ} as defined by (18) is based on the equality between the covariance matrices of the observed random

ural and controlled effects (see our footnote 14) coincide. This is called the no-interaction assumption, in which the mediation and direct effects do not depend on the values of the treatment T .

variables and unobserved error terms, namely

$$\mathbf{X} = \Psi \cdot \mathbf{X} + \epsilon \Rightarrow (\mathbf{I} - \Psi) \mathbf{X} = \epsilon \Rightarrow (\mathbf{I} - \Psi) \Sigma_{\mathbf{X}} (\mathbf{I} - \Psi)' = \Sigma_{\epsilon}. \quad (19)$$

Let $\tilde{\Sigma}_{\epsilon}$ be the covariance matrix (18) under our key assumption that $\rho_{TY} = 0$. This covariance matrix is displayed in (21).

$$\begin{aligned} \tilde{\Sigma}_{\mathbf{X}} &\equiv \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ -\beta_T^Z & 1 & 0 & 0 \\ 0 & -\beta_M^T & 1 & 0 \\ 0 & -\beta_Y^T & -\beta_Y^M & 1 \end{bmatrix}}_{\mathbf{I}-\Psi} \cdot \underbrace{\begin{bmatrix} \sigma_{ZZ} & \sigma_{ZT} & \sigma_{ZM} & \sigma_{ZY} \\ \cdot & \sigma_{TT} & \sigma_{TM} & \sigma_{TY} \\ \cdot & \cdot & \sigma_{MM} & \sigma_{MY} \\ \cdot & \cdot & \cdot & \sigma_{YY} \end{bmatrix}}_{\Sigma_{\mathbf{X}}} \cdot \underbrace{\begin{bmatrix} 1 & -\beta_T^Z & 0 & 0 \\ 0 & 1 & -\beta_M^T & -\beta_Y^T \\ 0 & 0 & 1 & -\beta_Y^M \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{(\mathbf{I}-\Psi)'} = \quad (20) \\ &= \underbrace{\begin{bmatrix} \sigma_{\epsilon_Z}^2 & 0 & 0 & 0 \\ \cdot & \sigma_{\epsilon_T}^2 & \rho_{TM} \sigma_{\epsilon_T} \sigma_{\epsilon_M} & 0 \\ \cdot & \cdot & \sigma_{\epsilon_M}^2 & \rho_{MY} \sigma_{\epsilon_M} \sigma_{\epsilon_Y} \\ \cdot & \cdot & \cdot & \sigma_{\epsilon_Y}^2 \end{bmatrix}}_{\Sigma_{\epsilon} \text{ under } \rho_{TY}=0} \equiv \tilde{\Sigma}_{\epsilon}. \quad (21) \end{aligned}$$

The equality (20)–(21) compares two covariance matrices of dimension four. Each matrix has $4 \cdot 4 = 16$ elements. We use $\tilde{\Sigma}_{\epsilon}[i, j]$ to denote the element in the i -th row and j -th column of matrix $\tilde{\Sigma}_{\epsilon}$. For sake of notational simplicity, we define $\tilde{\Sigma}_{\mathbf{X}} \equiv (\mathbf{I} - \Psi) \Sigma_{\mathbf{X}} (\mathbf{I} - \Psi)'$, where $\tilde{\Sigma}_{\mathbf{X}}[i, j]$ denotes the element in the i -th row and j -th column of the matrix $(\mathbf{I} - \Psi) \Sigma_{\mathbf{X}} (\mathbf{I} - \Psi)'$. The matrix is symmetric, thus the equality generates ten equations: four diagonal equations and six off-diagonal ones. Notationally, these ten equalities are defined by $\tilde{\Sigma}_{\mathbf{X}}[i, j] = \tilde{\Sigma}_{\epsilon}[i, j]$ for $i \leq j; i, j \in \{1, 2, 3, 4\}$. Four of the six off-diagonal equations are equal to zero, namely, $\tilde{\Sigma}_{\mathbf{X}}[1, j] = 0$ for $j \in \{2, 3, 4\}$ and $\tilde{\Sigma}_{\mathbf{X}}[2, 4] = 0$. What follows are the equations associated with the four zero elements in the covariance matrix (21):

$$\tilde{\Sigma}_{\mathbf{X}}[1, 2] = 0 \Rightarrow \sigma_{ZT} - \beta_T^Z \sigma_{ZZ} = 0 \Rightarrow \beta_T^Z = \frac{\sigma_{ZT}}{\sigma_{ZZ}}. \quad (22)$$

$$\tilde{\Sigma}_{\mathbf{X}}[1, 3] = 0 \Rightarrow \sigma_{ZM} - \beta_M^T \sigma_{ZT} = 0 \Rightarrow \beta_M^T = \frac{\sigma_{ZM}}{\sigma_{ZT}}. \quad (23)$$

$$\left. \begin{aligned} \tilde{\Sigma}_{\mathbf{X}}[1, 4] &= 0 \\ \tilde{\Sigma}_{\mathbf{X}}[2, 4] &= 0 \end{aligned} \right\} \Rightarrow \left. \begin{aligned} \sigma_{ZY} - \beta_Y^M \sigma_{ZM} - \beta_Y^T \sigma_{ZT} &= 0 \\ \sigma_{TY} - \beta_Y^M \sigma_{TM} - \beta_Y^T \sigma_{TT} &= 0 \end{aligned} \right\} \Rightarrow \left\{ \begin{aligned} \beta_Y^M &= \frac{\sigma_{ZT} \sigma_{TY} - \sigma_{TT} \sigma_{ZY}}{\sigma_{ZT} \sigma_{TM} - \sigma_{TT} \sigma_{ZM}} \\ \beta_Y^T &= -\frac{\sigma_{ZM} \sigma_{TY} - \sigma_{TM} \sigma_{ZY}}{\sigma_{ZT} \sigma_{TM} - \sigma_{TT} \sigma_{ZM}} \end{aligned} \right. \quad (24)$$

The four equalities in (22)–(24) suffice to identify all linear coefficients $\beta_T^Z, \beta_M^T, \beta_Y^M, \beta_Y^T$ of equations (12)–(15). There are six remaining equalities in (20)–(21). The four diagonal equations generated upon (20)–(21) identify the variances of the error terms. Those equations are listed in (25). The left-hand side of each equation consists of observed covariances or identified parameters. The right-hand side of each equation consists of the error variances.

$$\begin{aligned}
\tilde{\Sigma}_{\mathbf{X}}[1, 1] = \tilde{\Sigma}_{\mathbf{e}}[1, 1] &\Rightarrow \sigma_{ZZ} = \sigma_{\epsilon_Z}^2. \\
\tilde{\Sigma}_{\mathbf{X}}[2, 2] = \tilde{\Sigma}_{\mathbf{e}}[2, 2] &\Rightarrow (\sigma_{TT} - \beta_T^Z \sigma_{ZT}) - \beta_T^Z (\sigma_{ZT} - \beta_T^Z \sigma_{ZZ}) = \sigma_{\epsilon_T}^2. \\
\tilde{\Sigma}_{\mathbf{X}}[3, 3] = \tilde{\Sigma}_{\mathbf{e}}[3, 3] &\Rightarrow (\sigma_{MM} - \beta_M^T \sigma_{TM}) - \beta_M^T (\sigma_{TM} - \beta_M^T \sigma_{TT}) = \sigma_{\epsilon_M}^2. \\
\tilde{\Sigma}_{\mathbf{X}}[4, 4] = \tilde{\Sigma}_{\mathbf{e}}[4, 4] &\Rightarrow \begin{pmatrix} 1 \\ -\beta_Y^M \\ -\beta_Y^T \end{pmatrix}' \begin{bmatrix} \sigma_{YY} & \sigma_{MY} & \sigma_{TY} \\ \sigma_{MY} & \sigma_{MM} & \sigma_{TM} \\ \sigma_{TY} & \sigma_{TM} & \sigma_{TT} \end{bmatrix} \begin{pmatrix} 1 \\ -\beta_Y^M \\ -\beta_Y^T \end{pmatrix} = \sigma_{\epsilon_Y}^2.
\end{aligned} \tag{25}$$

The last two equalities can be extracted from (20)–(21) to identify the correlations ρ_{TM} and ρ_{MY} , which are described as follows:

$$\begin{aligned}
\tilde{\Sigma}_{\mathbf{X}}[2, 3] = \tilde{\Sigma}_{\mathbf{e}}[2, 3] &\Rightarrow \sigma_{TM} - \beta_T^Z \sigma_{ZM} - \beta_M^T (\sigma_{TT} - \beta_T^Z \sigma_{ZT}) = \rho_{TM} \sigma_{\epsilon_T} \sigma_{\epsilon_M}. \\
\tilde{\Sigma}_{\mathbf{X}}[3, 4] = \tilde{\Sigma}_{\mathbf{e}}[3, 4] &\Rightarrow \begin{pmatrix} 1 \\ -\beta_Y^M \\ -\beta_Y^T \end{pmatrix}' \begin{bmatrix} \sigma_{MY} & \sigma_{TY} \\ \sigma_{MM} & \sigma_{TM} \\ \sigma_{TM} & \sigma_{TT} \end{bmatrix} \begin{pmatrix} 1 \\ -\beta_M^T \end{pmatrix} = \rho_{MY} \sigma_{\epsilon_M} \sigma_{\epsilon_Y}.
\end{aligned} \tag{26}$$

With $\beta_T^Z, \beta_M^T, \beta_Y^M, \beta_Y^T$ identified by the four equalities in (22)–(24), and $\sigma_{\epsilon_Z}, \sigma_{\epsilon_T}, \sigma_{\epsilon_M}, \sigma_{\epsilon_Y}$ identified by the four equalities in (25), equation (26) is therefore just-identified for ρ_{TM} and ρ_{MY} .

The identification formulas to identify all linear coefficients $\beta_T^Z, \beta_M^T, \beta_Y^M, \beta_Y^T$ are described in the right-hand side of the expressions (22)–(24). Each identifying formula is associated with a well-known econometric estimator as follows:

1. Parameter β_T^Z is identified by (22) as the covariance between Z and T divided by the variance of Z . This formula implies that β_T^Z can be estimated by the OLS regression of T on Z .

$$\text{OLS: } T = \beta_T^Z \cdot Z + \epsilon_T. \tag{27}$$

2. Parameter β_M^T is identified by (23) as the ratio of the covariance between Z and M divided by the covariance of Z and T . This formula implies that β_M^T can be estimated by a 2SLS estimation where Z is the IV, T is the endogenous explanatory variable, and M is the outcome variable. Namely, β_M^T can be estimated by evaluating the standard 2SLS model as follows:

$$\text{First Stage: } T = \beta_T^Z \cdot Z + \epsilon_T, \quad (28)$$

$$\text{Second Stage: } M = \beta_M^T \cdot \hat{T} + \epsilon_M, \quad (29)$$

where $\hat{T}$ stands for the estimated values of T in the first stage.

3. Parameters β_Y^M, β_Y^T are jointly identified by the two remaining equations in (24). In [Online Appendix F](#) we show that β_Y^M and β_Y^T are the expected values of the estimators of a 2SLS regression where T plays the role of a conditioning variable, Z is the instrument, M is the endogenous variable, and Y is the dependent variable. Namely, β_Y^M and β_Y^T can be estimated by evaluating the following two-stage model:

$$\text{First Stage: } M = \gamma_M^Z \cdot Z + \gamma_M^T \cdot T + \epsilon_T, \quad (30)$$

$$\text{Second Stage: } Y = \beta_Y^M \cdot \hat{M} + \beta_Y^T \cdot T + \epsilon_Y, \quad (31)$$

where $\hat{M}$ are the estimated values of M in the first stage.

The estimation procedure associated with identification formulas (28) and (29) is the standard IV approach, and commonly understood. By contrast, the estimation procedure associated with identification formulas (30) and (31) is novel. In fact, it is a novel property of the framework laid out here that Z is a valid instrument to identify the causal effect of M on Y when conditioned on T . For this, and the intuition for the first stage (30), we refer the reader back to the exclusion restriction of Lemma [L-2](#), i.e. $Z \perp\!\!\!\perp Y(m)|T$, as well as Corollary [C-1](#).

Having established our core result and the intuition for our model, we also want to make the link to Table [1](#) explicit. In it, *Model I* stands for the standard IV model that evaluates the effect of T on M using Z as the instrument. This model is estimated by the 2SLS regression defined by equations (28) and (29). *Model III* is the mediation model with IV. This model is estimated by *Model I* plus the 2SLS regression represented by the linear equations (30)–(31). *Model II* stands for the IV model

that evaluates the total effect (TE) of T on Y directly. In Table 1, we have $T = f_T(Z, \epsilon_T)$ on $Y = g_Y(T, \eta_Y)$. The independence relation $Z \perp\!\!\!\perp (\epsilon_T, \eta_Y)$ induces the exclusion restriction $Y(t) \perp\!\!\!\perp Z$, and T is endogenous due to the statistical dependence between error terms ϵ_T and η_Y . *Model II* is obtained from *Model III* by substitution of the mediation variable M in (29) into the outcome equation in (31) as follows:

$$Y = \beta_Y^M \cdot M + \beta_Y^T \cdot T + \epsilon_Y \text{ and } M = \beta_M^T \cdot T + \epsilon_M \quad (32)$$

$$\Rightarrow Y = \beta_Y^M \cdot (\beta_M^T \cdot T + \epsilon_M) + \beta_Y^T \cdot T + \epsilon_Y \quad (33)$$

$$= \underbrace{(\beta_Y^M \cdot \beta_M^T + \beta_Y^T)}_{TE} \cdot T + \underbrace{\beta_Y^M \epsilon_M + \epsilon_Y}_{\eta_Y} \equiv g_Y(T, \eta_Y). \quad (34)$$

The error term η_Y of *Model II* is mapped into $\beta_Y^M \epsilon_M + \epsilon_Y$ in (34). Thus, the correlation between η_Y and ϵ_T results from the correlation between ϵ_M and ϵ_T , and the independence $Z \perp\!\!\!\perp (\epsilon_T, \eta_Y)$ is due to $Z \perp\!\!\!\perp (\epsilon_T, \epsilon_M, \epsilon_Y)$ in A-1.²²

Remark 2.5 In Online Appendix H, we extend the identification results and estimation methods of Section 2.3 to enable conditioning on additional covariates $\mathbf{K}$. The identification analysis amounts to simply replacing the covariance matrix $\tilde{\Sigma}_{\mathbf{X}}$ of the observed variables $\mathbf{X}$ by the covariance matrix $\tilde{\Sigma}_{\mathbf{X}|\mathbf{K}}$ of variables $\mathbf{X}$ conditioned on covariates $\mathbf{K}$. The identified parameters can be estimated by adding those covariates to the OLS regression (27) and the two 2SLS regressions defined in (28)–(29) and (30)–(31).

2.4 Allowing for General Error Dependency

As previously noted, Assumption A-2 is equivalent to assuming that T is endogenous in a regression of Y on T only because of confounders that affect T and M . Under linearity, Assumption A-2 implies $\rho_{TY} = 0$, and generates a model in which all parameters are point-identified. We view Assumption A-2 as quite plausible in our setting, but there are evidently many IV mediation settings (i.e. settings that can be described by equations (12)–(15)), where the identifying assumption is harder to defend. It is therefore helpful to extend the framework to relax Assumption A-2, i.e. allowing $\rho_{TY} \neq 0$. This is equivalent to leaving the statistical dependence among error terms entirely unrestricted. In Appendix C.1, we show how to generate a simulated dataset such that ρ_{TY}

²² In Online Appendix G, we investigate the particular case in which the instrument Z consists of a single variable. We show that the estimate of the total effect $TE = \beta_Y^M \cdot \beta_M^T + \beta_Y^T$ evaluated by the 2SLS regressions in (28)–(29) and (30)–(31) is numerically the same as the standard 2SLS estimate of the causal effect of T on Y in *Model II*, namely, the 2SLS estimate of θ_Y^T that uses $T = \beta_T^Z \cdot Z + \epsilon_T$ for the first stage and $Y = \theta_Y^T \cdot T + \epsilon_Y$ for the second stage.

can take any value.

Under $\rho_{TY} \neq 0$ we have a model with eleven instead of ten parameters: four coefficients $\beta_T^Z, \beta_M^T, \beta_Y^T, \beta_Y^M$, four error variances $\sigma_{\epsilon_Z}^2, \sigma_{\epsilon_T}^2, \sigma_{\epsilon_M}^2, \sigma_{\epsilon_Y}^2$, and three correlations $\rho_{TM}, \rho_{TY}, \rho_{MY}$. The identification of these coefficients relies on the matrix equality in (19), which yields ten identifying equalities. In [Appendix C.2](#), we show that these equalities render six of the eleven model parameters point-identified. Those are the coefficients β_T^Z, β_M^T , the correlation ρ_{TM} , and the variances $\sigma_{\epsilon_Z}^2, \sigma_{\epsilon_T}^2, \sigma_{\epsilon_M}^2$. For the remaining five parameters, $\beta_Y^M, \beta_Y^T, \rho_{TY}, \rho_{MY}, \sigma_{\epsilon_Y}^2$, we are left with four equalities. Since these five parameters are thus not point-identified, neither the direct effect $DE = \beta_Y^T$ nor the indirect effect $IE = \beta_M^T \cdot \beta_Y^M$ are point-identified. We can still evaluate bounds for these effects. In [Appendix C.2](#), we show a procedure for doing so in which we express parameters $\beta_Y^M, \beta_Y^T, \rho_{MY}$ and $\sigma_{\epsilon_Y}^2$ as functions of the values that parameter ρ_{TY} could take. In turn, we impose simple model restrictions on the covariance structure of the data to bound the possible range of ρ_{TY} . Figure 1 illustrates the bounds of the parameters when we generate a simulated dataset, as described in [Appendix C.1](#), with a true value of the correlation $\rho_{TY} = 0.6$ (solid vertical line). Each possible value of ρ_{TY} within the lower and upper bound (dashed vertical lines around 0.2 and 0.7) implies values of the other four parameters $\beta_Y^M, \beta_Y^T, \rho_{MY}$, and $\sigma_{\epsilon_Y}^2$. [Appendix C.1](#) generates the simulated data so that $\beta_T^Z = \beta_M^T = \beta_Y^T = \beta_Y^M = \sigma_{\epsilon_Y} = 1$. At the true value of $\rho_{TY} = 0.6$, the values of the β_Y^T, β_Y^M and σ_{ϵ_Y} also attain their true values, that is, $\sigma_{\epsilon_Y} = \beta_Y^T = \beta_Y^M = 1$. In [Section 5](#), we will apply this bounding exercise to our empirical application.

Figure 1: Bounds for Parameters $\rho_{TY}, \sigma_{\epsilon_Y}, \rho_{MY}, \beta_Y^T$ and β_Y^M .

Notes: This figure presents the bounds for model parameters $\rho_{TY}, \sigma_{\epsilon_Y}, \rho_{MY}, \beta_Y^T, \beta_Y^M$. The dashed vertical lines denote the bounds on ρ_{TY} . These in turn generate the bounds of all remaining parameters. The true value of ρ_{TY} , that is, $\rho_{TY} = 0.6$, is marked by a vertical solid line. Note that at the true value of $\rho_{TY} = 0.6$, the values of the β_Y^T, β_Y^M and σ_{ϵ_Y} also attain their true values, that is, $\sigma_{\epsilon_Y} = \beta_Y^T = \beta_Y^M = 1$.

3 Data

Our data is organized as a stacked panel of first differences between election dates, 1987 to 1998 and 1998 to 2009, staying as close as possible to the decadal changes usually studied in the literature. We study German regions' exposure to trade with Eastern Europe and China, which was exogenously affected by the fall of Communism and China's WTO accession. In Germany, imports from and exports to China and Eastern Europe roughly tripled from 1987 to 1998 (from about 20 billion to about 60 billion Euros each),²³ and again tripled between 1998 and 2009.

Our data is observed at the county (*Landkreis*) level.²⁴ We drop all city-states from the sample, and follow Dauth et al. (2014) in excluding East German counties from the first period of analysis, but including them in the second period. This imbalanced nature of the panel has no bearing on

²³Throughout the paper, we report values in thousands of constant-2005 Euros using exchange rates from the German Bundesbank.

²⁴ We follow Dauth et al. (2014) in using counties as a representation of German local labor markets. Dauth et al. (2014) show that results are qualitatively identical when using broader 'functional labor markets' but at the cost of econometric precision.

the estimation because we include period-specific region fixed effects for four broad regions in all estimations (North, West, South, and East Germany). Indeed, none of our results are affected at all by dropping East Germany altogether; we include it primarily to stay close to the existing literature. We observe 408 counties in our data, 86 of which are in East Germany. Over the two periods, we have 730 $((408 - 86) + 408)$ observations in total. For reference, we represent the data as two separate *Landkreise*-maps for the two periods in [Appendix D](#).

With a view toward the mediation framework we develop in [Section 2](#), we need the following variables: Treatment T_{it} is our measure of local labor market i 's import exposure in period t . Mediators M_{it} are labor market variables, and Final outcome Y_{it} refers to voting outcomes. Finally, we construct Z_{it} as an IV for T_{it} . We now explain how these variables are measured.²⁵

3.1 Import Exposure (Treatment T)

Following [Autor et al. \(2013\)](#) and [Dauth et al. \(2014\)](#), we construct *net import exposure* as follows.²⁶

$$T_{it} = \sum_j \frac{L_{ijt}}{L_{jt}} \frac{\Delta IM_{Gjt} - \Delta EX_{Gjt}}{L_{it}}. \quad (35)$$

ΔIM_{Gjt} denotes changes in Germany's imports in industry j in period t . Local labor market i 's composition of employment at the beginning of period t determines its exposure to changes in industry-specific trade flows ΔIM_{Gjt} over the ensuing decade.²⁷ Sector j receives more weight if region i 's national share of that sector $\frac{L_{ijt}}{L_{jt}}$ is high, but a lower weight if i 's overall workforce L_{it} is larger. [Autor et al. \(2013\)](#) focus on imports (ΔIM_{Gjt}) and consider the net of imports (ΔIM_{Gjt}) minus exports (ΔEX_{Gjt}) only in their appendix. [Dauth et al. \(2014\)](#) show that in Germany imports from and exports to low-wage manufacturing countries are not only more balanced in the aggregate than in the U.S. but also correlate positively at the industry level. As a result, when working with German data it is better to consider a local labor market's *net* import exposure, and this is what we do throughout the paper.

One concern with the measure of import exposure in equation (35) is that it is a composite effect

²⁵ Conditioning variables K_{it} are discussed with the results in [Section 4.1](#).

²⁶ Throughout the paper, we refer to net import exposure as import exposure for short.

²⁷ The *Institut für Arbeitsmarkt- und Berufsforschung* (IAB) reports industries of employment L_{ij} in standard international trade classification (SITC), and we link these to the U.N. Comtrade trade data using the crosswalk described in [Dauth et al. \(2014\)](#), which covers 157 manufacturing industries.

of the relative importance of trade-intensive industries *and* the relative importance of manufacturing employment in a region (i.e. $\frac{1}{L_{it}}$ relative to $\sum_j L_{ijt}$). The share of manufacturing employment might independently shape subsequent labor-market and voting changes. This well-known problem is solved by always conditioning on region i 's initial share of manufacturing employment in all our regressions (Autor et al., 2013).

3.2 Labor Market Variables (Mediator M)

We use the *Institut für Arbeitsmarkt- und Berufsforschung* (IAB)'s Historic Employment and Establishment Statistics database to glean information on workers' industry of employment, occupation, and place of work for all German workers subject to social insurance.²⁸ Individual-level data are aggregated up to the *Landkreis* level to match our voting data. We consider decadal changes in (i) total employment, (ii) manufacturing's employment share, (iii) manufacturing wages, (iv) non-manufacturing wages, and (v) unemployment, with data for the last one coming from the *German Statistical Office*. [Online Appendix I](#) provides additional information on data sources and variable construction.

3.3 Voting (Final Outcome Y)

To measure how import exposure affects voting behavior, we focus on party-votes in federal elections in Germany (*Bundestagswahlen*).²⁹ Due to its at-large voting system, Germany, like most continental European countries, has consistently had a multiparty system that spans the full political spectrum from far-left to extreme-right parties. This fact allows us to contrast the effect of import exposure on populists parties' vote share with that for moderate parties. We label four parties as "established" in that they were persistently represented in parliament over the 22 years we study. There is also many small parties whose vote share is always far below the 5 percent threshold of party votes needed to enter the federal parliament.³⁰ We collected these data to cre-

²⁸see Bender, Haas and Klose (2000) for a detailed description. Civil servants and self-employed individuals are not included in the data. Furthermore, we exclude workers younger than 18 or older than 65, and we exclude all individuals in training and in part-time jobs because their hourly wages cannot be assessed.

²⁹ The party vote, called *Zweitstimme*, mainly determines a party's share of parliamentary seats. German voters also cast a second vote for individual candidates, called *Erststimme*. This vote for individuals affects the very composition of party factions in the parliament, but it has no significant influence on their overall parliamentary share. Moreover, the decision on individual candidates might be strategic. We thus follow Falck et al. (2014) and focus on the party vote.

³⁰This threshold is not binding if a party wins at least three seats through the vote for individual candidates (*Erststimme*). During our period of analysis, this occurred once, in 1994. The individual candidates of the PDS party

ate a novel dataset of party vote shares at the county level. We group the small parties into three categories: far-left, extreme-right, other small parties. Altogether, *Landkreis*-level voting outcomes are divided into changes in the vote-share of (i) four mainstream parties (the CDU, the SPD, the FDP and the Green party), (ii) extreme-right parties, (iii) far-left parties, (iv) other small parties, and (v) turnout; see [Falck et al. \(2014\)](#).³¹

3.4 Others' Import Exposure (Instrument Z)

Endogeneity concerns in estimating the effect of import exposure on labor markets and voting come from the fact that domestic demand and supply shocks may simultaneously affect T_{it} , local labor market outcomes, and local voting behavior.

To overcome this problem, we follow the approach in [Autor et al. \(2013\)](#), instrumenting Germany's net imports from China and Eastern Europe, $\Delta IM_{Gjt} - EX_{Gjt}$, with the average imports from (exports to) China and Eastern Europe of a similar set of high-wage economies, ΔIM_{Ojt} (ΔEX_{Ojt}):³²

$$Z_{it}^{IM} = \sum_j \frac{L_{ijt-1}}{L_{jt-1}} \frac{\Delta IM_{Ojt}}{L_{it-1}}, \quad Z_{it}^{EX} = \sum_j \frac{L_{ijt-1}}{L_{jt-1}} \frac{\Delta EX_{Ojt}}{L_{it-1}}. \quad (36)$$

Finally, again following [Autor et al. \(2013\)](#), we lag the initial employment shares by one decade to address reverse causality concerns, denoting the lag by the subscript $t - 1$.

4 Baseline Results

In this section, we estimate *Model I* (trade effect on labor markets) and *Model II* (trade effect on voting) using the standard IV approach.

won four seats by *Erststimme*. As a result, the party received 30 seats in total, according its 4.4 percent of party votes (*Zweitstimme*) received.

³¹ [Online Appendix J](#) provides additional background on the German political system and party landscape, and in [Online Appendix K](#), we present descriptive patterns on the vote share variables by period.

³² We choose the same countries as [Dauth et al. \(2014\)](#) to instrument German imports and exports: Australia, Canada, Japan, Norway, New Zealand, Sweden, Singapore, and the United Kingdom. This set of countries excludes Eurozone countries because their demand conditions are likely correlated with Germany's.

4.1 Model I

Under linearity, *Model I* can be estimated by evaluating the following two-stage model:

$$T_{it} = \beta_T^Z \cdot Z_{it} + \beta_T^K \mathbf{K}_{it} + \eta_{it}, \quad (37)$$

$$M_{it} = \beta_M^T \cdot T_{it} + \beta_M^K \mathbf{K}_{it} + \epsilon_{it}. \quad (38)$$

Equations (37) and (38) are equations (28) and (29) with control variables $\mathbf{K}_{it}$ included and subscripts added to reflect the panel nature of the data. *Model II* will be estimated by the same model, with Y_{it} replacing M_{it} in the second stage (38).

$\mathbf{K}_{it}$ includes i 's start-of-period manufacturing employment share; the employment share in the largest sector;³³ employment share in the chemical industry;³⁴ the start-of-period employment share that is foreign-born; share of female population and share of population at retirement age; and the start-of-period unemployment rate; finally, start-of-period turnout and parties' vote-shares. Also included is a set of period-specific region fixed effects (North, West, and South); the regions are comparable to U.S. Census divisions (Dauth et al., 2014).³⁵ The regional fixed effects are period-specific to allow for different trends by period. Standard errors ϵ_{it} are clustered at the level of 96 larger economic zones defined by the Federal Office for Building and Regional Planning (BBR).

We use the exact same set of control variables in *Model I* and *Model II*. The core economic controls are well-motivated and routinely included in the related literature. Adding social and baseline voting controls is motivated by *Model II* and turns out to have more bearing on the estimates of interest in *Model II* than in *Model I*, as one would expect. We emphasize that *Model I* has been examined in detail in other papers, including Autor et al. (2013), whose results have already been replicated for Germany by Dauth et al. (2014). We therefore keep the results for *Model I* brief. There is, however, one aspect that we need to explore for the purposes of our analysis. There is naturally more than one observed labor market outcome that is likely to be impacted by import competition. In addition to (i) total employment, we observe (ii) manufacturing's employment

³³It is a feature of the German economy that some regions are dominated by one specific industry. In such regions, individual firms (e.g. Daimler-Benz, Volkswagen, or Bayer) are likely to have political bargaining power, and as a result politicians may help buffer trade shocks to limit adverse employment effects.

³⁴To account for the spatial clustering of this specific industry in Germany.

³⁵Each of West Germany's 10 states (*Bundesländer*) is fully contained inside one of these three regions.

share, (iii) manufacturing wages, (iv) non-manufacturing wages, and (v) unemployment. Total employment (i) is probably the most direct measure of the potential consequences of import exposure, but we find that (ii) and (iii) are also significantly impacted by import competition, with weak effects also for (iv) and (v).³⁶

Without additional separate dedicated sets of instruments for each potential mediator, our method can only identify the effect of as many mechanisms as there are treatments. Indeed, it is likely to be a common problem for researchers interested in applying the method developed here that there will often be a number of observed variables that potentially link a treatment T to an outcome Y . We therefore need to aggregate the multiple mechanisms into a single index. A *principal component* analysis is attractive in this regard because it generates indices that are purely statistical measurements based on the total variation in labor market outcomes and are orthogonal to one another by construction.³⁷ This approach is appealing as long as the mediating effects are sharply concentrated in a single principal component that has a clear interpretation. We label the principal components as “labor market components” (LMC).

Table 2: German Labor Markets’ Principal Components’ Factor-Loadings

	(1) $\Delta \log$ employment	(2) Δ Share Manuf. Empl.	(3) $\Delta \log$ Manuf. Wage	(4) $\Delta \log$ Non-Manuf. Wage	(5) Δ Share Unempl.
LMC_1	0.1711	-0.3632	0.5108	0.5486	0.5261
LMC_2	0.7625	0.6004	0.2104	0.0607	-0.1012
LMC_3	-0.5389	0.397	0.5311	0.3251	-0.4053

Notes: The table reports on the factor loadings of the five labor market variables on LMC_1 and LMC_2 . See discussion in the text. LMC_1 ’s eigenvalue is 2.707, explaining 54.1 percent of the total variation. LMC_2 ’s eigenvalue is 1.281, explaining 25.6 percent of the total variation. LMC_3 ’s eigenvalue is 0.509, explaining 10.2 percent of the total variation.

One can best interpret the LMCs through their relation to the labor market outcomes we observe, specifically through their factor loadings. Table 2 reports on the LMCs’ factor loadings. By construction, there are always as many LMCs as variables but to keep this section brief, we

³⁶ This is consistent with prior research that clearly shows that the labor market effects of import exposure are concentrated in manufacturing employment (Autor et al., 2013; Dauth et al., 2014).

³⁷ By contrast, methods that take weighted averages (Christensen and Miguel, 2016; Kling, Liebman and Katz, 2007) are usually applied to creating an outcome-index, but are unattractive for creating a mediating variable index precisely because they pre-impose weights. Similarly, *factor analysis* is more suitable when there are strong priors on how to group variables (Heckman, Pinto and Savelyev, 2013).

report only the first three. The convention is to regard only the LMCs with an eigenvalue larger than 1, i.e. only the first two in our case. In our data, LMC_1 and LMC_2 together explain about 80 percent of the variation in the labor market data ($0.541 + 0.256$). The third LMC explains only 10.2 percent, the fourth and fifth together explain the remaining 10 percent. LMC_1 is somewhat ambiguous: it has a weak positive correlation with overall employment but a negative one with manufacturing employment. Moreover, it is positively correlated with changes in wages, but also with unemployment.³⁸ By contrast, LMC_2 's interpretation is unambiguous: its factor loadings are strongly positive for changes in manufacturing employment and total employment, and negative for changes in unemployment. Thus, while LMC_1 appears to capture the divergence between manufacturing jobs and the rest of the labor market, LMC_2 captures common changes between manufacturing jobs and the overall labor market, i.e. the core labor market adjustments that are most affected by import exposure.

The estimation results of *Model I* are displayed in Table 3, each cell reporting on a different regression specification. In column 1, our least conservative specification considers only the start-of-period manufacturing employment share as a control.³⁹ In column 2, we account for the disproportionate regional employment share of some firms by including a control for the employment share in the largest sector, along with a control for the employment share in the chemical industry. In column 3, we add controls for the start-of-period employment share that is foreign-born, share of female population, population share at retirement age, and the start-of-period unemployment rate. In column 4, we add start-of-period turnout and party vote-shares. The addition of controls in columns 3 and 4 is motivated by *Model II*, but we will need to maintain the same control variables for labor market outcomes in our mediation framework and therefore apply the same here.

In the upper part of Panel A, we investigate the five individual labor market outcomes, where we view total employment as the primary one. Import exposure has a significant negative effect

³⁸ Our interpretation of LMC_1 is that it reflects the polarization of high-wage countries' labor markets (Goos, Manning and Salomons, 2009, 2014), associated with both higher wages and higher unemployment. A related view on LMC_1 is provided by the urban agglomeration literature, where Duranton and Puga (2005) point out that regional specialization has become "functional" as opposed to "sectoral" over the past decades, implying a tendency for headquarters and business services to cluster in large cities, a trend that appears to be clearly borne out in Germany (Bade, Laaser and Soltwedel, 2003).

³⁹ We always control for a region's start-of-period manufacturing share in employment, because it inherently drives part of the variation in T_{it} ; see the discussion in Section 3.4.

Table 3: *Model I*: Effect of Import Exposure T on Labor Markets M

	(1)	(2)	(3)	(4)
Panel A: Second Stage (38), for Individual Labor Market Outcomes				
$\Delta \log \text{employment}$	-0.023*** [0.004]	-0.024*** [0.006]	-0.024*** [0.003]	-0.022*** [0.006]
$\Delta \text{Share Manuf. Empl.}$	-0.440** [0.048]	-0.486** [0.018]	-0.684*** [0.000]	-0.732*** [0.000]
$\Delta \log \text{Manuf. Wage}$	-0.006** [0.013]	-0.006** [0.014]	-0.005** [0.020]	-0.005** [0.018]
$\Delta \log \text{Non-Manuf. Wage}$	-0.005*** [0.004]	-0.005*** [0.002]	-0.003** [0.046]	-0.002 [0.309]
$\Delta \text{Share Unempl.}$	0.076 [0.271]	0.061 [0.426]	0.127* [0.053]	0.082 [0.232]
LMC_1	-0.032 [0.495]	-0.032 [0.473]	0.024 [0.512]	0.029 [0.409]
LMC_2	-0.265*** [0.003]	-0.278*** [0.002]	-0.322*** [0.000]	-0.308*** [0.001]
LMC_3	-0.028 [0.523]	-0.028 [0.564]	-0.060* [0.097]	-0.056 [0.156]
LMC_4	0.023 [0.435]	0.011 [0.698]	-0.001 [0.966]	-0.035 [0.255]
LMC_5	0.017 [0.415]	0.011 [0.589]	0.007 [0.749]	-0.009 [0.691]
Panel B: First Stage Equation (37)				
β_M^{IM}	0.225*** [0.000]	0.212*** [0.000]	0.214*** [0.000]	0.217*** [0.000]
β_M^{EX}	-0.206*** [0.000]	-0.206*** [0.000]	-0.206*** [0.000]	-0.202*** [0.000]
Controls	Baseline	+ Industry	+Socio	+ Voting
R-Squared	0.524	0.551	0.555	0.566
F-Stat. of excluded Instruments	43.81	43.64	40.15	38.77
Period-by-region FE	Yes	Yes	Yes	Yes
Observations	730	730	730	730

Notes: (a) Each cell reports results from a separate IV regression. The data is a stacked panel of first-differences at the *Landkreis* level. Each regression has 730 observations, i.e. 322 *Landkreise* in West Germany, observed in the two periods 1987–1998 and 1998–2009, and 86 *Landkreise* in East Germany 1998–2009. We drop three city-states (Hamburg, Bremen, and Berlin in the East). (b) All specifications include region-by-period fixed effects. Columns incrementally add controls: Column 1 controls only for start-of-period manufacturing and period-specific region fixed effects. Column 2 adds further controls for industrial structure listed in the text. Column 3 adds further socio-economic and demographic controls listed in the text. Column 4 adds beginning-of-period voting controls. (c) Across rows, we investigate different outcomes. For example, the top row reports a semi-elasticity where a one-standard-deviation increase in T_{it} (€1,372 per worker) decreased total employment by about 3 percent, ($e^{-0.022 \cdot 1.372} - 1 = -0.03$). (d) The bottom panel reports the first-stage results. *p-values* are reported in square brackets, standard errors are clustered at the level of 96 commuting zones. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

on total employment, and on manufacturing employment and wages, as well as a weak effect on the unemployment rate. In our preferred specification in column 4, a one-standard-deviation increase in T_{it} (€1,372 per worker) decreases total employment by about 3 percent ($e^{-0.022 \cdot 1.372} - 1 = -0.03$).⁴⁰ In the lower part of Panel A, we investigate the effect of T_{it} on the five principal components LMC₁–LMC₅. We argued that PC analysis is appealing if the mediating effects turn out to be sharply concentrated in one PC, and this PC has a clear interpretation. Indeed, this turns out to be the case here: comparing the results for LMC₂ to those of the other four, it is the only one significantly impacted by import competition in all specifications, and the p-value is 0.001 or lower in all specifications. By contrast, the other four LMCs are largely unaffected by import competition.⁴¹ For the estimation of *Model III*, we will later focus on LMC₂ because it appears to summarize the effects of import exposure on labor markets very well in this data.

Panel B reports on the first-stage results of estimating equation (37). These results are highly significant, have the expected sign, and are in line with estimates in the existing literature.⁴² We do not study any individual-level labor market results; we refer the interested reader to [Dauth et al. \(2014\)](#), who present such evidence for Germany.

4.2 Model II

We now turn to estimating *Model II*, which shares the first-stage equation (37) with *Model I*, but replaces M_{it} with Y_{it} in the second-stage equation (38). We devote more attention to *Model II* because our empirical application is an investigation of the causes of political populism and the subsequent results are original to this study. In Table 1, in the Introduction, *Model II* represents causal equations $T = f_T(Z, \epsilon_T)$ and $Y = g_Y(T, \eta_Y)$. The endogeneity implied by $\epsilon_T \not\perp \eta_Y$ can be solved with the same instrument as in *Model I* because $Z \perp \epsilon_T, \eta_Y$.

Table 4 presents our baseline results of this 2SLS estimation. Each cell reports results from a different regression. While we are primarily interested in political support for populists, our voting data allow us to study effects on the entire political spectrum, i.e. changes in the vote-

⁴⁰ We report corresponding OLS results in [Online Appendix L](#) (Table [Online Appendix Table 4](#)).

⁴¹ Given our interpretation of the LMCs, these results resonate with existing evidence that import exposure has had large effects on (overall and manufacturing) employment, while the polarization of work and the rise of service jobs (i.e. our LMC₁) were explained by other factors, primarily automation ([Autor, Dorn and Hanson, 2015](#)).

⁴² For added clarity, we break the instrument $\beta_T^Z \cdot Z_{it}$ in the first-stage equation (37) into a separate import and export instrument $\beta_T^{IM} \cdot Z_{it}^{IM} + \beta_T^{EX} \cdot Z_{it}^{EX}$. Whether we have one instrument or a vector of instruments has no bearing on our argument that we develop a method that requires instruments dedicated to only one endogenous variable T .

shares of moderate, small, extreme-right, and far-left parties, as well as turnout. Each row in Table 4 pertains to one of these different outcome variables. Columns refer to different regression specifications, which are introduced in exactly the same manner as in Table 3. In addition, column 5 reports the results from our preferred specification in column 4 as standardized coefficients to facilitate a comparison of the magnitudes of the effect on different election outcomes.

The effects are broadly consistent across all specifications, though we see that the stepwise inclusion of controls reduces the estimated magnitude. Looking at the whole political spectrum, it is only the support of extreme-right parties that consistently and significantly responds to increasing import competition. We see no significant effects on turnout or on established, small, or far-left parties in our preferred specification in column 4 of table 4. Only the FDP vote share shows something close to a significant response to import exposure, but this response is much less significant than the response of the extreme-right parties' vote share and half the relative magnitude (in column 5), so that our focus is on the extreme-right parties.⁴³ In our preferred specification in column 4, a one-standard-deviation increase in T_{it} (€1,372 per worker) increases the extreme-right vote share by 0.13 ($0.096 \cdot 1.372$) percentage points, around 20 percent of the average per-decade increase of 0.64 percentage points during the 22 years we study. Column 5 reports the results from our preferred specification as beta coefficients, which shows that the estimated effects for all other parties are also economically insignificant as compared to the effect on extreme-right parties.⁴⁴

Interpreting the effects: Whether far-left or extreme-right populists capture protectionist sentiments is ultimately country-specific. As *The Economist* (July 30th, 2016) put it in a headline, "Farewell, left versus right. The contest that matters now is open against closed." Nonetheless, political scientists argue that the political left in Europe has found it difficult to take a coherent position against globalization over the past two decades, often hampered by internal intellectual conflicts (Arzheimer, 2009). Sommer (2008, p. 312) argues that "in opposing globalization, the left-wing usually criticizes an unjust and profit-oriented economic world order. [It] does not reject

⁴³ The less precisely estimated effect on the vote share of the market-liberal FDP may be interpreted as weak evidence for polarization. One possible explanation for this effect is that regions hit by a trade shock may face increasing demand for redistribution or government intervention in markets (Rodrik, 1995). As a result, those who do not approve such policies may choose to vote for the FDP. More generally, beneficiaries from import competition may demand increasing trade liberalization. Based on our reading of German politics, we take this as a hint for possible polarization, if the economically liberal FDP became an attractive choice for voters who position themselves against growing protectionist sentiments in their region.

⁴⁴ We present corresponding OLS estimates in Table Online Appendix Table 5 of Online Appendix L.

Table 4: *Model II*: Effect of Import Exposure (T) on Voting (Y)

	(1)	(2)	(3)	(4)	(5)
Δ Turnout	0.002 [0.348]	0.003 [0.266]	0.003 [0.225]	0.002 [0.254]	0.032 [0.254]
<u>Established Parties:</u>					
Δ voteshare CDU,CSU	-0.128 [0.457]	-0.159 [0.369]	-0.174 [0.310]	-0.090 [0.471]	-0.021 [0.471]
Δ voteshare SPD	-0.020 [0.897]	-0.004 [0.978]	0.024 [0.872]	0.026 [0.810]	0.002 [0.810]
Δ voteshare FDP	0.215*** [0.005]	0.204*** [0.009]	0.202*** [0.007]	0.120 [0.102]	0.022 [0.102]
Δ voteshare Greens	-0.132** [0.022]	-0.152*** [0.009]	-0.093* [0.079]	-0.056 [0.204]	-0.020 [0.204]
<u>Non-Established Parties:</u>					
Δ voteshare Extreme-Right	0.118*** [0.001]	0.129*** [0.001]	0.122*** [0.002]	0.096** [0.045]	0.047** [0.045]
Δ voteshare Far-Left	-0.037 [0.773]	-0.035 [0.793]	-0.087 [0.440]	-0.081 [0.434]	-0.021 [0.434]
Δ voteshare Other Small	-0.015 [0.696]	0.018 [0.640]	0.006 [0.874]	-0.015 [0.713]	-0.012 [0.713]
Controls	Baseline	+ Industry	+Socio	+ Voting	$\sim$ (4)
Period-by-region F.E.	Yes	Yes	Yes	Yes	Yes
Observations	730	730	730	730	730

Notes: (a) Each cell reports results from a separate instrumental variable regression. The data is a stacked panel of first-differences at the *Landkreis* level. Each regression has 730 observations, i.e. 322 *Landkreise* in West Germany, observed in the two periods 1987–1998 and 1998–2009, and 86 *Landkreise* in East Germany 1998–2009. We drop three city-states (Hamburg, Bremen, and Berlin). (b) All specifications include region-by-period fixed effects. Columns incrementally include added controls in identical fashion as Table 3. Column 4 is our preferred specification, which includes all controls. Column 5 reports on the same specification, with standardized outcome variables to facilitate comparison of magnitudes across rows. (c) Across rows, we report on different voting outcomes. Our focus will be on the most significant, and as column 5 shows also the most sizeable voting response, which occurs on the extreme right. (d) The first stage is identical to that in Table 3. p -values are reported in square brackets, standard errors are clustered at the level of 96 commuting zones. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

globalization per se but rather espouses a different sort of globalization. In contrast, the solutions proposed by the extreme right keep strictly to a national framework. The extreme right’s claim, therefore, that it is the only political force that opposes globalization fundamentally [...] rings true.”⁴⁵ In [Appendix D](#), we provide anecdotal evidence linking local import exposure to increasing support for the extreme right in two German regions.

While it is clear that import exposure has increased the extreme right’s vote share in our data, the overall effect is small. This is partly mechanical because in our setup the fixed effects absorb bigger shifts in voting behavior. More important, however, Germany did not have a populist party with broad appeal during our study period. All anti-globalization parties at the right fringe were extremist parties with neo-Nazi ties and associations to the *Third Reich*, which made them anathema to most Germans. Where populist leaders have broad appeal, the political backlash to import exposure may be more strongly reflected in changing vote shares. The coefficient size is thus specific to the political context. Our focus is, however, not on the magnitude of the effect of import exposure on voting behavior but on the causal mechanisms underlying it.⁴⁶

Robustness to Gravity Estimation: We also estimate results based on gravity residuals. This approach does not use IV but instead estimates the exogenous evolution of industry-specific Chinese and Eastern European comparative advantage over Germany based on a comparison of bilateral trade flows of Germany and ‘China plus Eastern Europe’ vis-a-vis a set of relevant destination markets.⁴⁷ The gravity results are reported in [Appendix E](#) and are in line with those in [Table 4](#). Reporting these gravity results is standard in the related literature ([Autor et al., 2013](#); [Dauth et al., 2014](#)), but our focus is naturally on the IV setting to which our identification framework applies.

Corroborating Evidence in an Individual-Level Analysis: One benefit of using German data is that Germany’s Socio-Economic Panel (SOEP) has a long-run panel structure that is unique

⁴⁵ For illustration, we excerpt from the extreme-right NPD’s “candidate manual”: “Globalization is a planetary spread of the capitalist economic system under the leadership of the Great Money. Despite by its very nature being Jewish-nomadic and homeless, it has its politically and militarily protected locus mainly on the East Coast of the United States” ([Grumke, 2012](#), p. 328).

⁴⁶ Aside from the size of estimated coefficients, we also note that Germany had relatively balanced trade with low-wage manufacturing countries during our study period and did not experience the “China Shock” in the same way as the U.S. and other high-wage countries. See [Online Appendix K](#).

⁴⁷ See [Autor et al. \(2013\)](#) and [Dauth et al. \(2014\)](#) for a discussion of the gravity-residuals approach relative to the IV approach.

among attitudinal socioeconomic surveys, starting in 1984 (GSOEP, 2007).⁴⁸ Importantly, we can locate individuals inside their local labor markets. As a result, we can associate individual workers w with their local labor market i 's import exposure (T), instrument T with Z as before, and add the same set of regional controls. This allows us to track decadal changes in individuals' party preferences in a way that mirrors our main local labor market analysis.⁴⁹ For our purpose, the relevant GSOEP question asks: "If there was an election today, who would you vote for?" We translate this question into a series of dummies that reflect the full party spectrum also observed in Table 4, e.g. one dummy if the individual would vote for the CDU, one if the individual would vote for the SPD, etc. The results, reported in Appendix F, mimic closely our main Table 4. A county's import exposure shifts individuals' preferences to the extreme right. Splitting the sample by worker types, we find the results to be entirely driven by low-skill workers, and more specifically those in manufacturing sectors, who are also most likely to experience adverse labor market effects from import exposure.

Without *Model III*, this is as far as we can go. Standard IV methods generate a causally identified effect of import exposure on total employment ($\beta_M^T = -0.027$) and a causally identified effect of import exposure on voting ($TE = 0.102$). We now estimate the mediation model to identify the extent to which the former explains the latter.

5 *Model III: Mediation Analysis*

In this section, we apply the estimation framework developed in Section 2 to estimate *Model III*. This allows us to estimate the indirect effect IE of import exposure on voting that runs through labor markets. The extent to which import exposure polarized voters because it caused labor market adjustments is identified by a comparison of this indirect effect with the total effect of import exposure on voting.⁵⁰

⁴⁸ In the U.S., the *General Social Survey* (GSS) for example added a panel component only in 2008.

⁴⁹ Because the SOEP only started to ask about voting intentions for the full party spectrum in 1990, we use the time windows 1990–1998 and 1998–2009, i.e. a slightly shorter period 1 compared to our main results.

⁵⁰ The total effect is estimated in *Model II* as 0.0963, reported in Table 4. The indirect effect is $\hat{\beta}_M^T$, reported in Table 3, multiplied by the effect of M on Y , which we now estimate.

Model III can be estimated by evaluating the following two-stage model:

$$M_{it} = \gamma_M^Z \cdot Z_{it} + \gamma_M^T \cdot T_{it} + \gamma_M^K \mathbf{K}_{it} + \epsilon_{it}, \quad (39)$$

$$Y_{it} = \beta_Y^M \cdot M_{it} + \beta_Y^T \cdot T_{it} + \beta_Y^K \mathbf{K}_{it} + \eta_{it}. \quad (40)$$

Equations (39) and (40) are exactly equations (30) and (31) in section 2.3, except that control variables are included and subscripts added to reflect the panel nature of the data.⁵¹

Which mediators should one consider in estimating equations (39) and (40)? Table 3 showed that two mediators in particular, i.e. the share of manufacturing employment as well as total employment, were most strongly impacted by import exposure. The principal component that summarizes the common variation in these two variables, LMC₂, was also very strongly impacted by import exposure. Since we cannot unpack the mediating effects of manufacturing employment and overall employment separately, we view LMC₂ as our key mechanism, as it is likely to capture all of the core labor market adjustments to import exposure. Nonetheless, for illustrative purposes, Table 5 reports the results of estimating equations (39) and (40) for all five labor market variables we have, as well as their first two principal components, across columns 1–7.

Panel A presents the results of the second-stage equation (40). As one may have expected, columns 1, 2 and 7 show that it is the labor market variables that most responded to import exposure in Table 3 that in turn appear to be the only ones to significantly impact extreme-right voting. The log of manufacturing wages is the only labor market outcome that responded to import exposure in Table 3 and is not estimated to have a significant mediation effect in Table 5. The point estimate $\hat{\beta}_Y^M$ in column 1 indicates that a one-percent drop in employment raises the change in the extreme right's vote share by 0.06 percentage points, i.e. 6.181/100, relative to an average per-decade increase of 0.64 percentage points during the 22 years we study.

The real importance of the point estimates $\hat{\beta}_Y^M$ lies in generating the indirect effect $IE = \hat{\beta}_Y^M \times \hat{\beta}_M^T$. This, together with other parameters of the mediation model is reported in Panel B of Table 5. The $\hat{\beta}_M^T$ of import exposure on log employment was estimated in Table 3 (as *Model I*), to be -0.0216 . The $\hat{\beta}_M^T$ of import exposure on the share of manufacturing employment was estimated in Table 3 to be -0.7317 . The $\hat{\beta}_M^T$ of import exposure on LMC₂ was estimated in Table 3 (as *Model I*) to be

⁵¹ The first-stage coefficients are denoted by γ 's instead of β 's because they do not correspond to parameters in the causal model represented by equations (12)–(15).

Table 5: Estimates of *Model III*, the Mediation Model

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mediating Variables:	$\Delta \log$ Employm.	Δ Share Manuf. Employm.	$\Delta \log$ Manuf. Wage	$\Delta \log$ Non-Manuf. Wage	Δ Share Unempl.	LMC_1	LMC_2	LMC_2
<i>Assumption A-2:</i>	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Panel A: Second Stage (40)								
β_Y^M	-6.181*** [0.000]	-0.183** [0.031]	-2.962 [0.188]	-63.684 [0.231]	-0.073 [0.805]	-2.345 [0.192]	-0.391*** [0.004]	$[-0.2227, -0.4193]$
$DE: \beta_Y^T$	-0.051 [0.106]	-0.099*** [0.008]	-0.019 [0.273]	-0.012 [0.854]	-0.002 [0.956]	0.120 [0.390]	-0.071** [0.012]	$[0.0087, -0.0835]$
Panel B: <i>Model III</i> Parameters								
<i>Model I:</i> β_M^T	-0.0216	-0.7317	-0.005	-0.002	0.082	0.029	-0.308	-0.308
<i>IE:</i> $\beta_M^T \cdot \beta_Y^M$	0.134	0.134					0.121	$[0.069, 0.129]$
<i>Model II:</i> TE	0.0963	0.0963					0.0963	0.0963
$S = IE/TE$	0.139	0.139					0.125	$[0.711, 1.339]$
Panel C: First Stage Equation (39)								
γ_M^{IM}	-0.005** [0.011]	-0.030 [0.473]	0.001** [0.024]	-0.001 [0.107]	-0.020 [0.192]	-0.003 [0.698]	-0.032* [0.094]	
γ_M^{EX}	0.006*** [0.000]	0.145*** [0.003]	0.003*** [0.000]	0.001 [0.306]	-0.017 [0.264]	0.013 [0.156]	0.082*** [0.000]	
γ_M^T	0.001 [0.864]	-0.354*** [0.007]	-0.002 [0.281]	0.001 [0.477]	0.130*** [0.001]	0.071*** [0.002]	-0.077 [0.152]	
Observations	730	730	730	730	730	730	730	
R-Squared	0.504	0.570	0.727	0.912	0.765	0.928	0.402	

Notes: (a) Columns 1–7 present the results of estimating *Model III* on the five labor-market outcomes, and the first two principal components Column 8 shows the results from the bounding exercise described in section 2.4. (b) Panel A presents the results of the second-stage equation (40). Panel B shows the main parameters of the mediation model. Panel C presents the results of the first-stage equation (39), where the estimated coefficients have the sign suggested by the logic discussed in section 2.2. (c) *p-values* are reported in square brackets in columns 1–7, with standard errors clustered at the level of 96 commuting zones. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

−0.308. The indirect effect IE is the product of $\hat{\beta}_Y^M \times \hat{\beta}_M^T$. The estimated IE of 0.134 happens to be identical to the third digit for log employment and the share of manufacturing employment, i.e. across columns 1–2. For LMC_2 , our preferred summary measure of the impact of import exposure on all labor market outcomes, it is very similar at 0.121. The TE estimated in Table 4 (as *Model II*) is constant across columns. Finally, the implied share of the total effect that is mediated by labor market adjustments is 139 percent ($0.134/0.096$) when focusing on the log of total employment or the share of manufacturing employment, and it 125 when using our preferred summary measure LMC_2 .

These results suggest that other channels linking import exposure to voting need to be politically moderating in the aggregate. Indeed, $\hat{\beta}_Y^T$ in Table 5 consistently has a negative sign.⁵² Another way to state this is that if the only effect of import exposure was to decrease employment, the political response would stronger than the one actually observed.

Panel C presents the results of the first-stage equation (39). It is useful to develop a prior about the coefficients’ expected signs before estimating equation (39), because the interpretation is not as straightforward as it is for the standard first-stage equation (37). The literature worries about domestic, industry-specific demand conditions as source of confounding bias. German industries that experience negative domestic demand shocks are expected to see fewer imports and declining employment. As discussed in section 2.2, this would suggests that, conditional on Germany’s imports T , higher imports from China into other countries will “cause” additional reductions in German employment by virtue of acting as a proxy for negative German demand. The first-stage coefficients have signs that are in line with this argument. In column 1 for instance, other countries’ imports from China worth €1,000 per worker reduce total employment by 0.5 percent ($\hat{\gamma}_M^{IM} = -0.005$), while the same amount in exports to China increases German employment by 0.6 percent ($\hat{\gamma}_M^{EX} = 0.006$).⁵³

Columns 8 reports on the bounds we derive when we allow $\rho_{TY} \neq 0$. The parameter β_M^T in *Model I*, as well as the directly estimated TE in *Model II* are still point-identified when $\rho_{TY} \neq 0$. By contrast, the parameters β_Y^M, β_Y^T can only be bounded, using restrictions on the covariance in

⁵² While we can only speculate, potentially moderating channels may follow from the association between import exposure and rich-countries’ task-upgrading (Becker and Muendler, 2015), and switching production towards more differentiated and higher-markup varieties (Holmes and Stevens, 2014). Of course, broadly speaking one may interpret these as labor market adjustments as well but they are not part of the available labor market aggregates usually studied.

⁵³ As before, we use two instruments Z_{it}^{IM}, Z_{it}^{EX} as IVs for one endogenous variable, namely T .

the data; see section 2.4. The covariance structure in the data allows the mediated share IE/TE of the effect of import exposure on extreme-right voting that is explained by LMC_2 to lie between 71 percent and 134 percent.

In summary, under the identifying assumption $\rho_{TY} = 0$, the results reported here suggest that changes in labor markets explain about 125 percent of the observed polarizing effect of import exposure on voters. The results thus suggest that the core labor market adjustments to import exposure combined lead to a polarizing effect on voters that is one quarter larger than the actual observed effect, which has to imply that other channels linking import exposure to voting are both moderating and important. When we allow $\rho_{TY} \neq 0$, this leaving the statistical dependence among error terms entirely unrestricted, the bounds generated by the model restrictions suggest that labor market adjustments explain at least 71 percent of the polarizing effect of import exposure on voters. Therefore, even at the lower bound the evidence clearly suggests that any effective remedy for the populist response to trade integration needs to focus on labor markets first and foremost.

6 Conclusion

A substantial body of recent evidence shows that in high-wage manufacturing countries such as Germany and the U.S., import exposure has had significant detrimental effects on the labor market outcomes of manufacturing workers. In this paper, we show that import exposure has also induced voters to turn to protectionist, populist, and nationalist policy agendas represented by Germany’s extreme right. Our focus here is to ask whether this effect of import exposure on voting for the extreme right is explained by (mediated by) import exposure’s effect on labor markets. There is good reason to believe it is: The aggregate effects coincide in the data and are mirrored in an individual-level analysis where those most prone to tilt toward the right are also those most vulnerable to the labor market consequences of import exposure.

In trying to answer this question, we face an empirical problem that is common to many research settings: We can use standard IV methods to causally identify the effect of a treatment (import competition T) on an outcome (labor market adjustments M), as well as causally identifying the effect of the same treatment on another outcome (voting for the extreme right Y). But if

the effect of T on Y is likely to work through M , then standard IV methods cannot estimate the extent to which it does.

This matters if we care about the mechanisms linking T and Y . To make headway, we develop a new methodology that allows us to perform the required *mediation analysis* in an IV setting. Applying our method, we find that the effect of import exposure that is mediated by labor market adjustments is larger than the total effect of import exposure on extreme-right voting, which in turn implies that other channels that link import exposure to voting are moderating in the aggregate. Our findings provide a first causal estimate of the importance of labor market adjustments in explaining the effect of import exposure on voting. The novel methodology we develop for this purpose may be useful in a broad range of empirical applications studying causal mechanisms in IV settings. While our identifying assumption plausibly holds in our setting, we caution that researchers need to critically evaluate its validity before applying it to other mediation-type IV settings.

References

- Arzheimer, K.**, "Contextual Factors and the Extreme Right Vote in Western Europe, 1980-2002," *American Journal of Political Science*, 2009, 53 (2), 259–275.
- Autor, David, David Dorn, and Gordon H Hanson**, "Untangling trade and technology: Evidence from local labour markets," *The Economic Journal*, 2015, 125 (584), 621–646.
- , —, and **Gordon Hanson**, "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," *American Economic Review*, 2013, 103 (6), 2121–68.
- , —, —, and **Kaveh Majlesi**, "Importing Political Polarization? The Electoral Consequences of Rising Trade Exposure," *NBER Working Paper*, 2016.
- Bade, Franz-Josef, Claus-Friedrich Laaser, and Rüdiger Soltwedel**, "Urban specialization in the internet age empirical findings for Germany, Processed," *Kiel Institute for World Economics*, 2003.
- Baron, Reuben M and David A Kenny**, "The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations.," *Journal of personality and social psychology*, 1986, 51 (6), 1173.
- Becker, Sascha O and Marc-Andreas Muendler**, "Trade and tasks: an exploration over three decades in Germany," *Economic Policy*, 2015, 30 (84), 589–641.
- Bender, Stefan, Anette Haas, and Christoph Klose**, "IAB Employment Subsample 1975-1995 Opportunities for Analysis Provided by the Anonymised Subsample," *IZA Discussion Paper 117*, 2000.

- Blundell, Richard and James Powell**, "Endogeneity in Nonparametric and Semiparametric Regression Models," in L. P. Hansen M. Dewatripont and S. J. Turnovsky, eds., *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, Vol. 2, Cambridge, UK: Cambridge University Press, 2003.
- **and —**, "Endogeneity in Semiparametric Binary Response Models," *Review of Economic Studies*, July 2004, 71 (3), 655–679.
- Che, Yi, Yi Lu, Justin R Pierce, Peter K Schott, and Zhigang Tao**, "Does Trade Liberalization with China Influence US Elections?," Technical Report, National Bureau of Economic Research 2016.
- Christensen, Garret S and Edward Miguel**, "Transparency, reproducibility, and the credibility of economics research," Technical Report, National Bureau of Economic Research 2016.
- Dahl, Christian M., Martin Huber, and Giovanni Mellace**, "It's Never Too LATE: A New Look at Local Average Treatment Effects with or Without Defiers," *Discussion Papers on Business and Economics, University of Southern Denmark*, 2/2017, 2017.
- Dauth, Wolfgang, Sebastian Findeisen, and Jens Suedekum**, "The Rise of the East and the Far East: German Labor Markets and Trade Integration," *Journal of European Economic Association*, 2014, 12 (6), 1643–1675.
- Dippel, Christian, Robert Gold, and Stephan Heblich**, "Globalization and its (Dis-) Content: Trade Shocks and Voting Behavior," *NBER Working Paper*, 2015, (w21812).
- Dix-Carneiro, Rafael and Brian K Kovak**, "Trade liberalization and regional dynamics," *American Economic Review*, 2017, 107 (10), 2908–46.
- Duranton, Gilles and Diego Puga**, "From sectoral to functional urban specialisation," *Journal of Urban Economics*, 2005, 57 (2), 343–370.
- Falck, Oliver, Robert Gold, and Stephan Heblich**, "E-lections: Voting Behavior and the Internet," *American Economic Review*, 2014, 104 (7), 2238–65.
- Feigenbaum, James J and Andrew B Hall**, "How Legislators Respond to Localized Economic Shocks: Evidence from Chinese Import Competition," *Journal of Politics*, 2015, 77 (4), 1012–30.
- Frölich, Markus and Martin Huber**, "Direct and indirect treatment effects: causal chains and mediation analysis with instrumental variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2017, pp. n/a–n/a.
- Goos, Maarten, Alan Manning, and Anna Salomons**, "Job polarization in Europe," *The American Economic Review*, 2009, 99 (2), 58–63.
- , — , **and —**, "Explaining job polarization: Routine-biased technological change and offshoring," *The American Economic Review*, 2014, 104 (8), 2509–2526.
- Grumke, Thomas**, *The Extreme Right in Europe*, Vandenhoeck & Ruprecht, 2012.
- GSOEP**, "The German Socio-Economic Panel Study (SOEP) - Scope, Evolution and Enhancements," Technical Report 1 2007.
- Hafeneger, Benno and Sven Schönfelder**, *Politische Strategien gegen die extreme Rechte in Parlamenten. Folgen für kommunale Politik und lokale Demokratie*, Friedrich-Ebert-Stiftung: Berlin, 2007.

- Heckman, James and Rodrigo Pinto**, "Unordered Monotonicity," *Forthcoming Econometrica*, 2017.
- Heckman, James J. and Edward J. Vytlacil**, "Structural Equations, Treatment Effects and Econometric Policy Evaluation," *Econometrica*, May 2005, 73 (3), 669–738.
- **and Rodrigo Pinto**, "Causal Analysis after Haavelmo," *Econometric Theory*, 2015, 31 (1), 115–151.
- Heckman, James J and Rodrigo Pinto**, "Econometric mediation analyses: Identifying the sources of treatment effects from experimentally estimated production technologies with unmeasured and mismeasured inputs," *Econometric reviews*, 2015, 34 (1-2), 6–31.
- Heckman, James J., Rodrigo Pinto, and Peter A. Savelyev**, "Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes," *American Economic Review*, 2013, 103 (6), 2052–2086.
- Holmes, Thomas J and John J Stevens**, "An alternative theory of the plant size distribution, with geography and intra-and international trade," *Journal of Political Economy*, 2014, 122 (2), 369–421.
- Imai, Kosuke, Luke Keele, and Dustin Tingley**, "A General Approach to Causal Mediation Analysis," *Psychological Methods*, 2010, 15 (4), 309–334.
- , — , **and Te Yamamoto**, "Identification, Inference and Sensitivity Analysis for Causal Mediation Effects," *Statistical Science*, 2010, 25 (1), 51–71.
- , — , **Dustin Tingley, and Teppei Yamamoto**, "Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies," *American Political Science Review*, 2011, 105 (4), 765–789.
- Imbens, Guido W. and Joshua D. Angrist**, "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, March 1994, 62 (2), 467–475.
- Jun, Sung Jae, Joris Pinkse, Haiqing Xu, and Neşe Yildiz**, "Multiple Discrete Endogenous Variables in Weakly-Separable Triangular Models," *Econometrics*, 2016, 4 (1).
- Kling, Jeffrey R, Jeffrey B Liebman, and Lawrence F Katz**, "Experimental analysis of neighborhood effects," *Econometrica*, 2007, 75 (1), 83–119.
- Krugman, Paul R**, "Trade and Wages, Reconsidered," *Brookings Papers on Economic Activity*, 2008, 2008 (1), 103–154.
- MacKinnon, David Peter**, *Introduction to statistical mediation analysis*, Routledge, 2008.
- Malgouyres, Clement**, "The Impact of Exposure to Low-Wage Country Competition on Votes for the Far-Right: Evidence from French Presidential Elections," *working paper*, 2014.
- Malgouyres, Clément**, "The impact of chinese import competition on the local structure of employment and wages: Evidence from france," *Journal of Regional Science*, 2017, 57 (3), 411–441.
- Mughan, Anthony and Dean Lacy**, "Economic Performance, Job Insecurity and Electoral Choice," *British Journal of Political Science*, 2002, 32 (3), 513–533.
- , **Clive Bean, and Ian McAllister**, "Economic globalization, job insecurity and the populist reaction," *Electoral Studies*, 2003, 22 (4), 617–633.

- New York Times**, “Ancient City’s Nazi Past Seeps Out After Stabbing,” 2009, *February 11th*.
- Pearl, Judea**, “The Mediation Formula: A Guide to the Assessment of Causal Pathways in Non-linear Models,” 2011. Forthcoming in *Causality: Statistical Perspectives and Applications*.
- , “Interpretation and identification of causal mediation,” *Psychological Methods, Special Section: Naturally Occurring Section on Causation Topics in Psychological Methods*, 2014, 19, 459–481.
- Pierce, Justin R and Peter K Schott**, “The Surprisingly Swift Decline of US Manufacturing Employment,” *American Economic Review*, 2016, 106 (7), 1632–62.
- Robins, James M. and Sander Greenland**, “Identifiability and Exchangeability for Direct and Indirect Effects,” *Epidemiology*, 1992, 3 (2), 143–155.
- Rodrik, Dani**, “Political economy of trade policy,” *Handbook of international economics*, 1995, 3 (3), 1457–1494.
- Sommer, Bernd**, “Anti-capitalism in the name of ethno-nationalism: ideological shifts on the German extreme right,” *Patterns of Prejudice*, 2008, 42 (3), 305–316.
- The Economist**, “The New Political Divide,” July 30th, 2016.

Appendix A Proofs of Exclusion Restrictions

Proof P-1 The treatment equation $T = f_T(Z, \epsilon_T)$ in equation (1) implies that $Z \not\perp\!\!\!\perp T$. Thus our task is to prove two exclusion restrictions: $Z \perp\!\!\!\perp M(t)$ and $Z \perp\!\!\!\perp Y(t)$. According to (4), the counterfactual mediation is given by $M(t) = f_M(t, \epsilon_M)$. But Assumption A-1 states that $Z \perp\!\!\!\perp (\epsilon_T, \epsilon_M, \epsilon_Y)$. In particular, we have that:

$$Z \perp\!\!\!\perp \epsilon_T \Rightarrow Z \perp\!\!\!\perp f_M(t, \epsilon_M) \Rightarrow Z \perp\!\!\!\perp M(t). \quad (41)$$

We can use iterated substitution to express the outcome counterfactual $Y(t)$ in equation (5) as the following function of error terms:

$$Y(t) = f_Y(t, M(t), \epsilon_Y) = f_Y(t, f_M(t, \epsilon_M), \epsilon_Y) \text{ by (4),} \quad (42)$$

$$\text{by A-1 we have that: } Z \perp\!\!\!\perp (\epsilon_M, \epsilon_Y) \Rightarrow Z \perp\!\!\!\perp f_Y(t, f_M(t, \epsilon_M), \epsilon_Y) \Rightarrow Z \perp\!\!\!\perp Y(t). \quad (43)$$

Proof P-2 The lemma requires two proofs. The first shows that Z is not independent of M conditioned on T , that is, $Z \not\perp\!\!\!\perp M|T$. The second shows that the exclusion restriction $Z \perp\!\!\!\perp Y(m)|T$ holds under the independence condition $\epsilon_T \perp\!\!\!\perp \epsilon_Y$ in Assumption A-2.

An intuitive justification for $Z \not\perp\!\!\!\perp M|T$ relies on interpreting the correlations generated by conditioning on T . Recall that the treatment equation is given by $T = f_T(Z, \epsilon_T)$. Thus, conditioning on $T = t$ is equivalent to conditioning on the values of Z, ϵ_T such that $f_T(Z, \epsilon_T) = t$. This induces a correlation between Z and ϵ_T and thereby $Z \not\perp\!\!\!\perp \epsilon_T|T$. Moreover, ϵ_T correlates with ϵ_M and therefore we also have that $Z \not\perp\!\!\!\perp \epsilon_M|T$. But if $Z \not\perp\!\!\!\perp \epsilon_M|T$, then $Z \not\perp\!\!\!\perp f_M(T, \epsilon_M)|T$ and therefore we have that $Z \not\perp\!\!\!\perp M|T$ as $M = f_M(T, \epsilon_M)$. In summary, conditioning on T induces a correlation between Z and ϵ_T , but error term ϵ_T correlates with ϵ_M , which in turn generates a correlation between Z and M . It remains to be shown that the independence relation $\epsilon_T \perp\!\!\!\perp \epsilon_Y$ generates the exclusion restriction $Z \perp\!\!\!\perp Y(m)|T$, where the outcome counterfactual $Y(m)$ is given by $Y(m) = f_Y(T, m, \epsilon_Y)$ as in equation (6). The following rationale justify this assessment. Assumptions A-1–A-2 generates the unconditional independence relation $(\epsilon_T, \epsilon_Z) \perp\!\!\!\perp \epsilon_Y$. Let $f_1(\cdot), f_2(\cdot), f_3(\cdot)$ be three arbitrary non-degenerate functions such that $f_1 : \text{supp}(\epsilon_Z) \times \text{supp}(\epsilon_T) \rightarrow \mathbb{R}$,

$f_2 : \text{supp}(\epsilon_Z) \rightarrow \mathbb{R}$, $f_3 : \text{supp}(\epsilon_Y) \rightarrow \mathbb{R}$. Under this notation, we have that:

$$(\epsilon_T, \epsilon_Z) \perp\!\!\!\perp \epsilon_Y \Rightarrow \epsilon_Z \perp\!\!\!\perp \epsilon_Y | f_1(\epsilon_Z, \epsilon_T) \Rightarrow f_2(\epsilon_Z) \perp\!\!\!\perp f_3(\epsilon_Y) | f_1(\epsilon_Z, \epsilon_T). \quad (44)$$

In particular, we can set functions $f_1(\epsilon_T)$, $f_2(\epsilon_Y)$, $f_3(\epsilon_Z, \epsilon_T)$ in (44) to the following expressions: $f_1(\epsilon_Z) = f_Z(\epsilon_Z)$, $f_2(\epsilon_Y) = f_Y(t, m, \epsilon_Y)$, and $f_3(\epsilon_Z, \epsilon_T) = f_T(f_Z(\epsilon_T), \epsilon_Z)$. Thus:

$$f_2(\epsilon_Z) \perp\!\!\!\perp f_3(\epsilon_Y) | f_1(\epsilon_Z, \epsilon_T) \quad (45)$$

$$\Rightarrow f_Z(\epsilon_Z) \perp\!\!\!\perp f_Y(t, m, \epsilon_Y) | (f_T(f_Z(\epsilon_T), \epsilon_Z) = t) \quad \forall (t, m) \in \text{supp}(T) \times \text{supp}(M) \quad (46)$$

$$\Rightarrow Z \perp\!\!\!\perp f_Y(t, m, \epsilon_Y) | (T = t) \quad \forall (t, m) \in \text{supp}(T) \times \text{supp}(M) \quad (47)$$

$$\Rightarrow Z \perp\!\!\!\perp Y(m) | T. \quad (48)$$

Proof P-3 Assumptions **A-1–A-2** implies that $\epsilon_Y \perp\!\!\!\perp (Z, \epsilon_T)$. According to equation (6), we have that:

$$\begin{aligned} P(Y(m) \leq y | T = t) &= P(f_Y(t, m, \epsilon_Y) \leq y | T = t), \\ &= P(f_Y(t, m, \epsilon_Y) \leq y | f_T(Z, \epsilon_T) = t), \\ &= P(f_Y(t, m, \epsilon_Y) \leq y), \\ &= P(Y(t, m) \leq y), \end{aligned}$$

where the third equality comes from $\epsilon_Y \perp\!\!\!\perp (Z, \epsilon_T)$.

Appendix B Generating the Error Structure and in Simulated Data

It is instructive to show how the dependence relations in Assumption A-2 translate into ρ_{TM} , ρ_{MY} , and ρ_{TY} in Σ_ϵ , equation (18). A simulated dataset with the dependence relations in A-2 can be straightforwardly generated in the following way:

- Separately generate error terms ϵ_T, ϵ_Y that are normally distributed with mean zero and variance one, $N(0, 1)$. These are statistically independent, i.e. $\epsilon_T \perp\!\!\!\perp \epsilon_Y$.
- Let error term ϵ_M be defined as $\epsilon_M = \sqrt{\omega} \cdot \epsilon_T + \sqrt{(1-\omega)} \cdot \epsilon_Y$ for any $\omega \in [0, 1]$.⁵⁴

The correlation between ϵ_M, ϵ_T is given by $\rho_{TM} = \sqrt{\omega}$. Thereby $\epsilon_M \not\perp\!\!\!\perp \epsilon_T$. By symmetry, we also have that $\rho_{MY} = \sqrt{(1-\omega)}$ and $\epsilon_M \not\perp\!\!\!\perp \epsilon_Y$. Having drawn ϵ_T, ϵ_Y independently implies that the correlation between ϵ_T and ϵ_Y is $\rho_{TY} = 0$. However, conditioning on $\epsilon_M = e$ induces a linear relation between ϵ_T, ϵ_Y , namely, $\epsilon_T = e/\sqrt{\omega} - \sqrt{(1-\omega)/\omega} \cdot \epsilon_Y$. Thus, the correlation between ϵ_T, ϵ_Y conditioned on ϵ_M is $\rho_{TY|\epsilon_M} = -1$ and thereby $\epsilon_T \not\perp\!\!\!\perp \epsilon_Y | \epsilon_M$. Figure A1 displays the model correlations $\rho_{TM}, \rho_{MY}, \rho_{TY}$ and $\rho_{TY|\epsilon_M}$ as a function of $\omega \in [0, 1]$. A high ω implies a high ρ_{TM} , i.e. pronounced bias in the relation between trade and labor markets. By contrast, a low ω implies a high ρ_{MY} , i.e. pronounced bias in the relation between labor markets and voting.

Notes: This figure presents the correlations among error terms $\epsilon_T, \epsilon_M, \epsilon_Y$. Properties of error terms are: (i) normally distributed with mean zero and variance one; (ii) $\epsilon_T \perp\!\!\!\perp \epsilon_Y$ and (iii) $\epsilon_M = \sqrt{\omega} \cdot \epsilon_T + \sqrt{(1-\omega)} \cdot \epsilon_Y$ where $\omega \in [0, 1]$. Parameters $\rho_{TM}, \rho_{MY}, \rho_{TY}$ stand for the correlations between (ϵ_T, ϵ_M) , (ϵ_M, ϵ_Y) , and (ϵ_T, ϵ_Y) respectively. Parameter $\rho_{TY|\epsilon_M}$ stands for the correlation between ϵ_T, ϵ_Y conditioned on error term ϵ_M .

⁵⁴ Note that $\epsilon_T \sim N(0, 1)$ and $\epsilon_Y \sim N(0, 1)$ imply $\epsilon_M \sim N(0, 1)$.

It is instructive to investigate the bias generated by a misspecified model in which T, M are assumed to be exogenous, i.e. in which the mutual independence of $\epsilon_T, \epsilon_M, \epsilon_Y$ is wrongly assumed. Let the data be generated by equations (12)–(15), and the model coefficients be normalized to equal 1, that is, $\beta_T^Z = \beta_M^T = \beta_Y^T = \beta_Y^M = 1$. The true parameters β_M^T, β_Y^T and β_Y^M are identified through equations (23)–(24). If the error terms $\epsilon_T, \epsilon_Y, \epsilon_M$ were wrongly assumed to be statistically independent, parameters $\beta_M^T, \beta_Y^T, \beta_Y^M$ could be estimated by OLS through the following equations:

$$\text{OLS: } \beta_M^T = \frac{\sigma_{TM}}{\sigma_{TY}}, \quad (49)$$

$$\text{OLS: } \beta_Y^T = \frac{\sigma_{MM}\sigma_{TY} - \sigma_{TM}\sigma_{MY}}{\sigma_{MM}\sigma_{TT} - \sigma_{TM}^2}, \quad (50)$$

$$\text{OLS: } \beta_Y^M = \frac{-\sigma_{TM}\sigma_{TY} + \sigma_{TT}\sigma_{MY}}{\sigma_{MM}\sigma_{TT} - \sigma_{TM}^2}. \quad (51)$$

Figure A2 displays the correct model parameters – evaluated by the 2SLS in (28)–(31) – and the OLS estimators identified by equations (49)–(51). While the true parameters are set to be 1, the OLS estimators may range from 0 to 2 depending on the error correlations. Since a high ω implies pronounced bias in the relation between trade and labor markets (a high ρ_{TM}), the OLS estimate of β_M^T diverges from the true value 1 as ω increases. By contrast, the OLS estimates of β_Y^T and β_Y^M converges to the true value 1.

Figure A2: Model Parameters by ω

Notes: The figure presents the model parameters $\beta_M^T, \beta_Y^T, \beta_Y^M$ computed under the right assumption that error correlate and under the mistaken assumption of no error correlation. If errors correlate according to A-2, then parameters β_M^T, β_Y^T and β_Y^M are identified by equations (23)–(24). Under the (wrong) assumption of no error correlation, the model parameters are identified by (49)–(51).

Appendix C Examining the Bounds Using Simulated Data

For the purposes of section 2.4, it is instructive to simulate a model that describes the basic features of the estimation method under correlated error terms. Appendix C.1 defines the equations that generate the correlated error terms, and simulates the data. Appendix C.2 describes a procedure for how $\rho_{TY}(\kappa)$ can be identified as a function of κ , and how the other four parameters β_Y^M , β_Y^T , ρ_{MY} , $\sigma_{\epsilon_Y}^2$ can be in turn identified.

Appendix C.1 A Structure for Correlated Error Terms

The error terms are generated on the basis of a parameter $\omega \in [0, 1]$ and four random variables $\xi_T, \xi_Y, \xi_M, \xi_E$ that are i.i.d. normally distributed with mean zero and variance 1, $N(0, 1)$. The error structure is defined by the following equations:

$$\epsilon_T = \sqrt{\omega} \cdot \xi_E + \sqrt{1 - \omega} \cdot \xi_T; \quad (52)$$

$$\epsilon_Y = \sqrt{\omega} \cdot \xi_E + \sqrt{1 - \omega} \cdot \xi_Y; \quad (53)$$

$$\epsilon_M = \sqrt{0.5} \cdot \xi_M + \sqrt{0.5} \cdot (\sqrt{0.5} \cdot \epsilon_T + \sqrt{0.5} \cdot \epsilon_Y); \quad (54)$$

Note that the correlation between error variables ϵ_T, ϵ_Y is given by $\rho_{TY} = \omega \in [0, 1]$. Equation (54) of error term ϵ_M is symmetric regarding errors ϵ_T and ϵ_Y . As a consequence, the correlations between ϵ_M, ϵ_T and ϵ_M, ϵ_Y are the same, that is, $\rho_{TM} = \rho_{MY}$. These correlations also depend on parameter ω . For instance, when $\omega = 0$ implies that $\rho_{TM} = \rho_{MY} = 0.5$ and $\omega = 1$ implies that $\rho_{TM} = \rho_{MY} = 0.815$. Figure A3 displays the relation between correlations $\rho_{YT}, \rho_{TM}, \rho_{MY}$ and parameter ω .⁵⁵

⁵⁵ Equations (52)–(54) also imply that error terms $\epsilon_T, \epsilon_M, \epsilon_Y$ are normally distributed with mean zero and variance 1. Finally, error term ϵ_Z is set to be normally distributed with mean zero and variance 1 and to be statistically independent of error terms $\epsilon_T, \epsilon_M, \epsilon_Y$.

Figure A3: Correlations Among Error Terms by ω

Notes: This figure presents the correlations among error terms

$$\begin{aligned}\epsilon_T &= \sqrt{\omega} \cdot \xi_E + \sqrt{1-\omega} \cdot \xi_T; \\ \epsilon_Y &= \sqrt{\omega} \cdot \xi_E + \sqrt{1-\omega} \cdot \xi_Y; \\ \epsilon_M &= \sqrt{0.5} \cdot \xi_M + \sqrt{0.5} \cdot (\sqrt{0.5} \cdot \epsilon_T + \sqrt{0.5} \cdot \epsilon_Y);\end{aligned}$$

where $\omega \in [0, 1]$ and $\xi_T, \xi_Y, \xi_M, \xi_E$ are i.i.d normally distributed random variables with mean zero and variance 1.

Appendix C.2 Bounding the Model Parameters

The subsequent identification follows the same steps utilized in in section 2.3. Model identification relies on the matrix equation $\tilde{\Sigma}_{\mathbf{X}} = \Sigma_{\epsilon}$, where $\tilde{\Sigma}_{\mathbf{X}} \equiv (\mathbf{I} - \Psi) \Sigma_{\mathbf{X}} (\mathbf{I} - \Psi)'$, as in (19). This yields ten linear equalities given by $\tilde{\Sigma}_{\mathbf{X}}[i, j] = \Sigma_{\epsilon}[i, j]$ for $i \leq j; i, j \in \{1, 2, 3, 4\}$. The independence relation $Z \perp\!\!\!\perp (\epsilon_T, \epsilon_M)$ implies that $\Sigma_{\epsilon}[1, 2] = \Sigma_{\epsilon}[1, 3] = 0$. Thereby, the equalities $\tilde{\Sigma}_{\mathbf{X}}[1, 2] = 0$ and $\tilde{\Sigma}_{\mathbf{X}}[1, 3] = 0$ as in (22)–(23) still hold. As a consequence, the coefficients β_T^Z, β_M^T remain unchanged and are still identified by $\beta_T^Z = \frac{\sigma_{ZZT}}{\sigma_{ZZ}}$ and $\beta_M^T = \frac{\sigma_{ZTM}}{\sigma_{ZT}}$. The coefficients β_T^Z and β_M^T can still be evaluated by the OLS regression in (27) and the 2SLS in (28)–(29), respectively. The coefficients β_T^Z, β_M^T refer to Model I in Table 1. The model is not altered by the causal relation between T and Y , and therefore β_T^Z, β_M^T are not affected by relaxing $\rho_{TY} \neq 0$.

Error variances are identified by the diagonal of the matrix equality $\tilde{\Sigma}_{\mathbf{X}} = \Sigma_{\epsilon}$. The equality $\tilde{\Sigma}_{\mathbf{X}}[1, 1] = \Sigma_{\epsilon}[1, 1]$ implies that $\sigma_{ZZ} = \sigma_{\epsilon_Z}^2$. Furthermore, the identification of β_T^Z and the observed

covariances enable the identification of error variance $\sigma_{\epsilon_T}^2$ by the following equation:

$$\tilde{\Sigma}_{\mathbf{X}}[2, 2] = \Sigma_{\mathbf{e}}[2, 2] \Rightarrow (\sigma_{TT} - \beta_T^Z \sigma_{ZT}) - \beta_T^Z (\sigma_{ZT} - \beta_T^Z \sigma_{ZZ}) = \sigma_{\epsilon_T}^2. \quad (55)$$

In addition, the identification of β_M^T enables the identification of $\sigma_{\epsilon_M}^2$ by the following equation:

$$\tilde{\Sigma}_{\mathbf{X}}[3, 3] = \Sigma_{\mathbf{e}}[3, 3] \Rightarrow (\sigma_{MM} - \beta_M^T \sigma_{TM}) - \beta_M^T (\sigma_{TM} - \beta_M^T \sigma_{TT}) = \sigma_{\epsilon_M}^2. \quad (56)$$

The parameters β_T^Z, β_M^T and variances $\sigma_{\epsilon_T}^2, \sigma_{\epsilon_M}^2$ enable the identification of correlation ρ_{TM} via the following equation:

$$\tilde{\Sigma}_{\mathbf{X}}[2, 3] = \Sigma_{\mathbf{e}}[2, 3] \Rightarrow \sigma_{TM} - \beta_T^Z \sigma_{ZM} - \beta_M^T (\sigma_{TT} - \beta_T^Z \sigma_{ZT}) = \rho_{TM} \sigma_{\epsilon_T} \sigma_{\epsilon_M}. \quad (57)$$

We are left with four equalities, $\tilde{\Sigma}_{\mathbf{X}}[i, j] = \Sigma_{\mathbf{e}}[i, j]$ such that $(i, j) \in \{(1, 4), (2, 4), (4, 4), (3, 4)\}$, and five parameters, $\beta_Y^M, \beta_Y^T, \rho_{TY}, \rho_{MY}, \sigma_{\epsilon_Y}^2$, that are not point-identified.

The best approach to examine the identification of the five model parameters that are not point-identified is to express all five parameters $\beta_Y^M, \beta_Y^T, \rho_{TY}, \rho_{MY}, \sigma_{\epsilon_Y}^2$, in terms of the product $\rho_{TY} \cdot \sigma_{\epsilon_Y}$, which we label as an auxiliary variable $\kappa \equiv \rho_{TY} \cdot \sigma_{\epsilon_Y}$. If κ were known, we could identify all model parameters. Specifically, let $\beta_Y^M(\kappa), \beta_Y^T(\kappa), \rho_{TY}(\kappa), \rho_{MY}(\kappa), \sigma_{\epsilon_Y}^2(\kappa)$ denote the values that the model parameters would take for a given value of κ . Let $\beta_Y^M(\kappa), \beta_Y^T(\kappa), \rho_{TY}(\kappa), \rho_{MY}(\kappa), \sigma_{\epsilon_Y}^2(\kappa)$ denote the values that the model parameters would take for a given value of κ . These parameters can be identified by the following procedure:

1. The equalities $\tilde{\Sigma}_{\mathbf{X}}[1, 4] = \Sigma_{\mathbf{e}}[1, 4]$, and $\tilde{\Sigma}_{\mathbf{X}}[2, 4] = \Sigma_{\mathbf{e}}[2, 4]$ generate the following equations:

$$\sigma_{ZY} - \beta_Y^M \sigma_{ZM} - \beta_Y^T \sigma_{ZT} = 0, \quad (58)$$

$$\sigma_{TY} - \beta_Y^M \sigma_{TM} - \beta_Y^T \sigma_{TT} = \underbrace{\rho_{TY} \sigma_{\epsilon_Y}}_{\kappa} \sigma_{\epsilon_T}. \quad (59)$$

Given a value of κ , the coefficients $\beta_Y^M(\kappa), \beta_Y^T(\kappa)$ can be obtained by the following formula:

$$\begin{bmatrix} \beta_Y^M(\kappa) \\ \beta_Y^T(\kappa) \end{bmatrix} = (\mathbf{B}' \mathbf{A}^{-1} \mathbf{B})^{-1} \cdot (\mathbf{B}' \mathbf{A}^{-1} \mathbf{C}), \quad (60)$$

$$\text{where } \mathbf{A} = \begin{bmatrix} \sigma_{ZZ} & \sigma_{ZT} \\ \sigma'_{ZT} & \sigma_{TT} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \sigma_{ZM} & \sigma_{ZT} \\ \sigma'_{TM} & \sigma_{TT} \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} \sigma_{ZY} \\ \sigma_{TY} - \kappa \cdot \sigma_{\epsilon_T} \end{bmatrix}. \quad (61)$$

2. The parameters $\beta_Y^M(\kappa), \beta_Y^T(\kappa)$ enable the evaluation of error variance $\sigma_{\epsilon_Y}^2(\kappa)$ via equality $\tilde{\Sigma}_{\mathbf{X}}[4, 4] = \Sigma_{\mathbf{e}}[4, 4]$ in (25). Namely:

$$\sigma_{\epsilon_Y}^2(\kappa) = \begin{pmatrix} 1 \\ -\beta_Y^M(\kappa) \\ -\beta_Y^T(\kappa) \end{pmatrix}' \begin{bmatrix} \sigma_{YY} & \sigma_{MY} & \sigma_{TY} \\ \sigma_{MY} & \sigma_{MM} & \sigma_{TM} \\ \sigma_{TY} & \sigma_{TM} & \sigma_{TT} \end{bmatrix} \begin{pmatrix} 1 \\ -\beta_Y^M(\kappa) \\ -\beta_Y^T(\kappa) \end{pmatrix}. \quad (62)$$

3. The evaluation of $\sigma_{\epsilon_Y}(\kappa)$ in addition to $\beta_Y^T(\kappa), \beta_Y^M(\kappa)$ as well as $\beta_M^T, \sigma_{\epsilon_M}$ enables the identification of the correlation $\rho_{MY}(\kappa)$ via the equality $\tilde{\Sigma}_{\mathbf{X}}[3, 4] = \Sigma_{\mathbf{e}}[3, 4]$ in (26). Namely:

$$\rho_{MY}(\kappa) = \begin{pmatrix} 1 \\ -\beta_Y^M(\kappa) \\ -\beta_Y^T(\kappa) \end{pmatrix}' \begin{bmatrix} \sigma_{MY} & \sigma_{TY} \\ \sigma_{MM} & \sigma_{TM} \\ \sigma_{TM} & \sigma_{TT} \end{bmatrix} \begin{pmatrix} (\sigma_{\epsilon_M} \sigma_{\epsilon_Y}(\kappa))^{-1} \\ -\beta_M^T \\ \frac{\sigma_{\epsilon_M}(\kappa) \sigma_{\epsilon_Y}(\kappa)}{\sigma_{\epsilon_M}(\kappa) \sigma_{\epsilon_Y}(\kappa)} \end{pmatrix}. \quad (63)$$

4. Finally, the correlation $\rho_{TY}(\kappa)$ can be identified by the ratio $\rho_{TY}(\kappa) = \kappa / \sigma_{\epsilon_Y}(\kappa)$.

The identification of model parameters is thus anchored on the variable κ . The variable κ^* is unknown, but we can bound it on the interval $\kappa \in [\kappa_0, \kappa_1]$ by imposing some simple model restrictions on the the covariance structure of the data: (i) Parameter $\rho_{TY}(\kappa)$ stands for the correlation between ϵ_T and ϵ_Y . Thus, we can delimit the values of κ such that $|\rho_{TY}(\kappa)| \leq 1$. Likewise, parameter $\rho_{MY}(\kappa)$ stands for the correlation between ϵ_M and ϵ_Y , such that $|\rho_{MY}(\kappa)| \leq 1$ must hold. (ii) Let $\Sigma_{\mathbf{e}}(\kappa)$ denote the identified error covariance matrix for a value of κ . $\Sigma_{\mathbf{e}}(\kappa)$ must be positive-definite matrix, that is to say that its eigenvalues must be strictly positive. Thus the values of κ must be such that the root-solutions of the λ -polynomial generated by the determinant $\det(\Sigma_{\mathbf{e}}(\kappa) - \lambda \mathbf{I}) = 0$ are strictly positive.⁵⁶ (iii) The outcome error variance $\sigma_{\epsilon_Y}^2$ is smaller than or equal to the variance of the outcome σ_{YY} itself, i.e. $\sigma_{YY} \geq \sigma_{\epsilon_Y}^2$. Because κ is defined as $\kappa \equiv \rho_{TY} \cdot \sigma_{\epsilon_Y}$, it must therefore lie in the interval $\kappa \in [-\sqrt{\sigma_{\epsilon_Y}}, \sqrt{\sigma_{\epsilon_Y}}]$. (iv) Lastly, we can add the restriction $0 \leq \rho_{MY}(\kappa) \leq 1$ on the model correlation.

We evaluate the interval $[\kappa_0, \kappa_1]$ that complies with all these model restrictions. With this interval, we can bound ρ_{TY} , and then bound the other four parameters as functions of ρ_{TY} . We can thus generate the bounds for the direct effect $DE = \beta_Y^T$; the indirect effect $IE = \beta_M^T \cdot \beta_Y^M$; the total effect $TE = \beta_Y^T + \beta_M^T \cdot \beta_Y^M$; the share of the total effect that is mediated by the indirect effect.

⁵⁶In our notation, $\mathbf{I}$ stands for the identity matrix of dimension 4.

Figure A4: T_{it} in 1987–1998 (Left), and 1998–2009 (Right)

Notes: Trade shocks mapped into 322 West German counties for 1987–1998 (left) and into 408 German counties for 1998–2009 (right). The two circles enclose the regions in Palatine (on the left) and Bavaria (on the right).

Appendix D Graphical Representation and Qualitative Evidence

Figure A4 shows the spatial dispersion of our key regressor T_{it} . While our empirical analysis will use fixed effects to wash out any secular trends, it is reassuring that even in the raw data there appears to be little correlation in T_{it} in space or time.⁵⁷ This reflects both Germany's diverse pattern of industrial production and the fact that the dominant driver of T_{it} changed from Eastern Europe in 1987–1998 to China in 1998–2009 (Dauth et al., 2014). The enclosed region in the southwest of our map is Southwest-Palatine (*Südwestpfalz*), a region that traditionally produced shoe and leather products. In our data, Southwest-Palatine is in the top decile of negatively shocked districts in both periods. In 2005, the region's biggest city, Pirmasens, had an unemployment rate of 20 percent. At the same time, extreme-right parties increased their vote-share from 1.3 percent in 1987 to 3.45 percent in 2009. A study commissioned by the Friedrich Ebert Foundation conducted interviews with local politicians which suggested that the local *Republikaner* managed to mobilize

⁵⁷We use period-specific fixed effects for four broad regions. With East Germany as one of the regions we thus have a total of seven period-by-region fixed effects in our stacked panel.

enough voter support to enter Pirmasens' city parliament by explicitly linking import competition to social hardships (Hafeneger and Schönfelder, 2007). The enclosed counties in the southeastern part of the map are Rottal-Inn, Passau, Freyung-Grafenau, Regen, and Cham. This is a traditional manufacturing region specialized in glass products and furniture making. This region too saw high import exposure, declining employment, and increasing support for the extreme right in the past decades, attracting international attention when neo-Nazis carried out a near-fatal attack on Passau's police chief in 2008 (New York Times, 2009). Additional descriptive statistics are reported in Online Appendix K.

Appendix E Identification of T on Y with Gravity Residuals

An alternative to the IV approach pursued in the paper is to estimate gravity equations, which exploit essentially the same source of exogenous variation. The endogeneity concern with increasing imports ΔIM_{Gjt} is that they do not only reflect increasing competitiveness of Chinese and Eastern European ('CE') industries,⁵⁸ but also German industry-specific demand changes. The gravity approach to solving this problem is to compare changes in German industries' exports to other countries O in relation to Chinese and Eastern European exports to O . This comparison reflects changes in Chinese and Eastern European comparative advantage over Germany, and allows constructing an exogenous measure ΔIM_{Gjt}^{grav} to replace ΔIM_{Gjt} .⁵⁹ Online Appendix M shows how to obtain the gravity-residuals ΔIM_{Gjt}^{grav} that replace ΔIM_{Gjt} and, similarly, the gravity-residuals ΔEX_{Gjt}^{grav} that replace ΔEX_{Gjt} . An exogenous measure for changes in in German industries' import exposure can be obtained from netting out both effects such that $\Delta Trade_{Gjt}^{grav} = \Delta IM_{Gjt}^{grav} - \Delta EX_{Gjt}^{grav}$. Substituting $\Delta IM_{Gjt} - EX_{Gjt}$ in equation (35) with $\Delta Trade_{Gjt}^{grav}$ provides an exogenous measure of regional import exposure based on the gravity approach as

$$T_{it}^{grav} = \sum_j \frac{L_{ijt}}{L_{jt}} \frac{\Delta Trade_{Gjt}^{grav}}{L_{it}} \quad (64)$$

We now substitute T_{it} from our baseline regression with T_{it}^{grav} directly. Otherwise, we run exactly the same specifications as before.

⁵⁸Competitiveness increases due to productivity increases, better market access, and decreasing relative trade cost.

⁵⁹We chose Belgium, France, Greece, Italy, Luxembourg, the Netherlands, Portugal, Spain, and the UK as "other countries" O for our gravity regressions, to be comparable with Dauth et al. (2014).

Table A1: Gravity Results for Effect of T_{it} on Voting

	(1) Baseline Gravity	(2) + Structure Gravity	(3) + Industry Gravity	(4) + Voting Gravity	(5) +Socio Gravity	(6) Standard. Gravity
Δ Turnout	0.000** (2.143)	0.000* (1.774)	0.000** (1.980)	0.000** (1.966)	0.000* (1.706)	0.002* (1.706)
<i><u>Established Parties:</u></i>						
Δ Vote Share CDU/CSU	0.006 (0.873)	0.003 (0.392)	0.003 (0.342)	0.001 (0.186)	0.001 (0.089)	0.000 (0.089)
Δ Vote Share SPD	-0.008 (-1.197)	-0.004 (-0.674)	-0.005 (-0.720)	0.004 (0.732)	0.004 (0.668)	0.000 (0.668)
Δ Vote Share FDP	0.001 (0.204)	0.006 (1.365)	0.004 (1.064)	0.007* (1.840)	0.007* (1.933)	0.001* (1.933)
Δ Vote Share Green Party	0.006** (2.021)	0.000 (0.071)	0.001 (0.408)	0.000 (0.047)	0.000 (0.012)	0.000 (0.012)
<i><u>Non-established Parties</u></i>						
Δ Vote Share Extreme-Right Parties	0.004* (1.855)	0.006** (2.430)	0.006** (2.276)	0.003 (1.575)	0.003* (1.779)	0.002* (1.779)
Δ Vote Share Far-Left Parties	-0.011* (-1.814)	-0.012* (-1.884)	-0.012* (-1.894)	-0.014** (-2.382)	-0.013** (-2.177)	-0.003** (-2.177)
Δ Vote Share Other Small Parties	0.002 (0.755)	0.002 (0.696)	0.003 (1.141)	-0.001 (-0.710)	-0.001 (-0.766)	-0.001 (-0.766)
Period-by-region F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Observations	730	730	730	730	730	730

Notes: (a) Each cell reports results from a separate regression. The data are a stacked panel of first-differences at the *Landkreis* level. Each regression has 730 observations, i.e. 322 *Landkreise* in West Germany in 1987–1998 and 1998–2009, plus 86 in Easter Germany in 1998–2009. We drop the three city-states (Hamburg, Bremen, and Berlin). (b) All specifications include region-by-period fixed effects. Column 1 controls only for start-of-period manufacturing. Column 2 adds controls for dominant industries (employment share of the largest industry, and chemicals). Column 3 adds controls for the structure of the workforce (share foreign workers, population share female, retirement age, and unemployed). Column 4 adds beginning-of-period voting controls. Finally, Column 5 presents specification (4) with standardized outcome variables to facilitate comparison. (c) All standard errors are clustered at the level of 96 commuting zones. All specifications include region-by-period fixed effects. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Results are reported in Table A1. Again, each cell reports results from a different regression. Rows specify different outcome variables, and columns refer to different regression specifications. Results are consistent with our main specifications reported in Table 4. The key observation is that the positive effect of import exposure on extreme-right party votes is confirmed by this alternative identification strategy. In addition, a few other effects that were insignificant in table 4 become sharper here: there is more evidence for a positive effect of import exposure on turnout, and the negative effect on far-left parties turns significant in the gravity regressions. Conversely, the positive effect on the vote share of the market-liberal party FDP turns insignificant. Overall, these patterns are all internally consistent, and do not detract from our focus on extreme-right vote shares in the IV setting.

Appendix F Individual-Level Evidence

In this section, we test whether our regional-level results can be confirmed at the individual level. For each party P , we aggregate individuals' self-reported voting intentions into a decadal cumulative share of years in which a respondent stated he would vote for it. We calculate Y_{wt}^P as the ratio of the number of years that w states a preference for party P , divided by the number of years that w answered the question in the SOEP.⁶⁰ For each party P , the dependent variable is a share between 0 and 1 for individual w in time period t and we separately estimate

$$Y_{wt}^P = \gamma_{Y-1}^Y \cdot Y_{wt-1}^P + \gamma_T^Y \cdot T_{it} + \gamma_X^Y \cdot X_{it-1} + \epsilon_{wt}. \quad (65)$$

for each party outcome. With a slight abuse of notation, Y_{wt-1}^P controls for w 's survey response to the same question in the base year. X_{it} refers to the same set of regional controls for the base-year as in table 4. Our focus is on estimating γ_T^Y , the effect of region i 's import exposure T_{it} on a resident worker w 's reported party support. Table A2 reports the results. Across rows it mimics closely our main table 4, except that there is no turnout measure in the SOEP. Every coefficient in table A2 reports the estimate of γ_T^Y from a separate regression. T_{it} is always instrumented as

⁶⁰ It is better to measure the outcome as a cumulative share for the whole period instead of using a first difference approach because the latter relies only on individuals' answer at the beginning and the end of the period. Moreover, respondents do not answer all questions in every year, which increases the number of missing observations in a first difference specification. As a result, we obtain about three times as many 'person-decade' observations, and correspondingly more precise estimates, using the share measure than with the first-difference measure.

before, we do not report the first stage regressions again. Column 1 includes the period-region fixed effects and the regional economic controls from table 4. We also add region i 's base-year socioeconomic and voting controls X_{it-1} from table 4 for each period. To better gauge magnitudes, column 2 reports the same specification with standardized outcomes. Import exposure shifts individuals' preferences to the extreme right, though the effect is weaker in the SOEP than it was in the actual voting data in table 4, with a t -statistic of only 1.62. By contrast there is stronger evidence of a reduction in preference for the established left party, the SPD. No other party across the spectrum shows a response that is close to being significant. We also find distinctive results on what types of workers are driving these patterns. In columns 5–7 we split the sample by skill as well as by whether an individual works in manufacturing, i.e. whether their employment sector is more heavily exposed to trade competition.⁶¹ Both the extreme right effect and the SPD effect are entirely driven by low-skill workers, while high-skill workers do not respond at all.⁶² Splitting the low-skill sample into manufacturing and non-manufacturing employment, we see that the extreme-right response is entirely driven by low-skill workers in manufacturing sectors. This implies that those who are most likely to experience adverse labor market effects from trade are the ones most likely to turn towards the extreme right because of increasing import exposure in their region. By contrast, the reduction in the change in the SPD's vote share is driven by low-skill *non*-manufacturing workers. A possible explanation is that low-skill workers in the service sector are affected by competition from laid-off manufacturing workers, or that laid-off manufacturing workers had to accept unattractive jobs in the service sector. In either case, they might blame the SPD-initiated labor market reforms.

⁶¹ In an earlier and related paper, we focused on comparing the effect of individuals' import exposure due to their industry of employment relative to their regions' import exposure (Dippel et al., 2015). However, we have since concluded that individuals' industry of employment is measured too coarsely in the SOEP to draw strong conclusions about the relative importance of these two types of import exposure.

⁶² The SOEP reports skills as educational attainment according to the ISCED-1997 classification, where 'high' means some college.

Table A2: Individual-Level Analysis

	(1)	(2)	(3)	(4)	(5)
	All Controls	Standardized	High-Skill	Low-Skill & Manuf.	Low-Skill & Not Manuf.
<i><u>Established Parties:</u></i>					
Would Vote CDU/CSU	0.001 (0.292)	0.003 (0.292)	-0.007 (0.278)	-0.013 (0.794)	0.008 (0.827)
Would Vote SPD	-0.008* (1.901)	-0.016* (1.901)	-0.013 (0.460)	-0.011 (0.400)	-0.017* (1.930)
Would Vote FDP	0.001 (0.459)	0.005 (0.459)	-0.018 (0.420)	0.011 (0.664)	0.007 (0.568)
Would Vote Green Party	0.003 (1.000)	0.012 (1.000)	0.070 (1.474)	0.025 (0.909)	0.002 (0.152)
<i><u>Non-Established Parties:</u></i>					
Would Vote Extreme-Right Parties	0.003 (1.619)	0.023 (1.619)	0.010 (0.875)	0.083** (2.206)	0.006 (0.475)
Would Vote Far-Left Parties	-0.001 (1.059)	-0.007 (1.059)	-0.051 (1.358)	0.019 (1.356)	-0.009 (1.055)
Would Vote Other Small Parties	-0.001 (0.642)	-0.007 (0.642)	0.005 (0.182)	-0.026 (1.053)	-0.003 (0.190)
Period-by-region F.E.	Yes	Yes	Yes	Yes	Yes
Observations	9,669	9,669	1,348	2,199	6,122

Notes: (a) Each cell in this table reports on a separate regression. An observation is an individual w over a period t , where we consider 1990–1998, and 1998–2009, closely mirroring the local labor market results. Each row reports on stated preferences for a different party. The outcome in each row is the ratio of the number of years that w states a preference for party P , divided by the number of years that w answered the question in the SOEP. The reported coefficient in all cells is the IV coefficient of regional import exposure T_{it} . (b) Column 1 is the baseline specification which includes period and four region fixed effects as well as all the regional economic, voting and demographic controls from table 4, and individuals' base-year stated political preferences. This is the full set of controls included in all columns. To better gauge magnitudes, column 2–5 standardize all outcomes by their mean. In columns 3–5, we break the sample by individuals' skill as well as by whether they are employed in the manufacturing sector ($1,348 + 2,199 + 6,122 = 9,669$). High-skill workers (column 3) do not change their political support at all in response to import exposure. Column 4 shows that it is the population most affected by import exposure – low-skill manufacturing workers – that drives the effects on the far right. (c) t -statistics are reported in round brackets, standard errors are clustered at the region level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Online Appendix

to

**“Instrumental Variables and Causal Mechanisms:
Unpacking the Effect of Trade on Workers and Voters”**

Online Appendix A The Mediation Model with No Confounders

It is illustrate how mediation works in the setting above in a simple randomized control trial (RCT), as in the seminal work of [Robins and Greenland \(1992\)](#), and using the language of counterfactual variables. In a mediation model, the causal effect of T on Y decomposes into the *direct*, and the *indirect* effects. The total effect TE stands for the average causal effect of T on Y . The direct effect DE stands for the causal effect of T on Y that is not generated by changes in M . The indirect effect IE is the causal effect of T on Y induced by the change in the distribution of the mediator M . Let the treatment assignment take values in $\text{supp}(T) = \{t_0, t_1\}$, where t_0 indicates the control group and t_1 indicates the treatment group. Let $F_{M(t)}(m)$ denote the cumulative density function (CDF) of the counterfactual mediator $M(t)$ conditional on the assignment $t \in \{t_0, t_1\}$. The total effect TE is the expected difference between counterfactual outcome Y when T is fixed at t_1 and t_0 . The direct effect $DE(t)$ evaluates the expected difference of counterfactual outcomes between treated (t_1) and control (t_0) group holding the distribution of the mediator fixed at $M(t)$. The indirect effect $IE(t)$ evaluates the expected value of the the difference between counterfactual outcomes $Y(t, m)$ when the distribution of the mediator m varies between treated $M(t_1)$ and control $M(t_0)$ while holding the t -input fixed. Formally, the other three causal effects (TE, DE, IE) are defined as follows:

$$\begin{aligned} TE &= E(Y(t_1) - Y(t_0)) && \equiv E(Y(t_1, M(t_1)) - Y(t_0, M(t_0))) \\ DE(t) &= E(Y(t_1, M(t)) - Y(t_0, M(t))) && \equiv \int E(Y(t_1, m) - Y(t_0, m)) dF_{M(t)}(m) \\ IE(t) &= E(Y(t, M(t_1)) - Y(t, M(t_0))) && \equiv \int E(Y(t, m)) [dF_{M(t_1)}(m) - dF_{M(t_0)}(m)] \end{aligned}$$

[Robins and Greenland's \(1992\)](#) main contribution is to show that the total effect of T on Y can be decomposed as the sum of the effect of T on Y that is mediated by M (the indirect effect) and the causal effect of T on Y that is not mediated by M (the direct effect).⁶³ Equations (66)–(67) express the total effect as the sum of direct and indirect effects:⁶⁴

$$\begin{aligned} TE &= E(Y(t_1, M(t_1)) - Y_i(t_0, M(t_0))) \\ &= \left(E(Y(t_1, M(t_1))) - E(Y(t_0, M(t_1))) \right) + \left(E(Y(t_0, M(t_1)) - Y_i(t_0, M(t_0))) \right) = DE(t_1) + IE(t_0) \end{aligned} \tag{66}$$

$$= \left(E(Y(t_1, M(t_1))) - E(Y(t_1, M(t_0))) \right) + \left(E(Y(t_1, M(t_0)) - Y_i(t_0, M(t_0))) \right) = IE(t_1) + DE(t_0). \tag{67}$$

[Robins and Greenland's \(1992\)](#) decomposition can be extended to the problem we examine by allowing T to be a continuous variable, in which case the decomposition is obtained by the total differentiation of the counterfactual outcome:

⁶³ [Pearl \(2011\)](#) makes a distinction between natural (or “descriptive”) direct and indirect effects and controlled (or “prescriptive”) direct effects.

⁶⁴ A large literature on mediation analysis relies on the Sequential Ignorability Assumption [A-3](#) of [Imai et al. \(2010\)](#) to identify mediation effects. We discuss this assumption in [Online Appendix B](#). See [Frölich and Huber \(2017\)](#) for a recent review of the mediation literature.

$$\underbrace{\frac{dE(Y(t))}{dt}}_{\text{Total Effect}} = \underbrace{\frac{\partial E(Y(t, m))}{\partial t}}_{\text{Direct Effect}} + \underbrace{\frac{\partial E(Y(t, m))}{\partial m} \cdot \frac{dE(M(t))}{dt}}_{\text{Indirect Effect}}. \quad (68)$$

Identification of the total, direct, and indirect effects hinges on the dependence relation among the error terms $\epsilon_T, \epsilon_M, \epsilon_Y$ in (1)–(3). Suppose that the error terms ϵ_T and ϵ_M are statistically independent, i.e. $\epsilon_T \perp\!\!\!\perp \epsilon_M$. This means there are no unobserved variables that jointly cause T and M . In this case, T is exogenous with respect to M , as in an RCT. It is easy to show that the independence condition $M(t) \perp\!\!\!\perp T$ holds and the expected value of counterfactual variable $M(t)$ is identified by the conditional expectation $E(M(t)) = E(M|T = t)$. In addition, if error terms (ϵ_T, ϵ_M) and ϵ_Y were statistically independent, then the independence conditions $(Y(t), Y(t, m)) \perp\!\!\!\perp T$ and $Y(t, m) \perp\!\!\!\perp M$ would also hold. This means that variables T and M are exogenous with respect to Y and that the expected value of counterfactual variables $E(Y(t))$ and $E(Y(t, m))$ would be identified by conditional expectations of observed variables $E(Y(t)) = E(Y|T = t)$ and $E(Y(t, m)) = E(Y|T = t, M = m)$, respectively.

Online Appendix B The Sequential Ignorability Assumption

A large literature on mediation analysis relies on the Sequential Ignorability Assumption [A-3](#) of [Imai et al. \(2010\)](#) to identify mediation effects.

Assumption A-3 *Sequential Ignorability* ([Imai et al., 2010](#)):

$$(Y(t', m), M(t)) \perp\!\!\!\perp T|X \quad (69)$$

$$Y(t', m) \perp\!\!\!\perp M(t)|(T, X), \quad (70)$$

where X denotes pre-intervention variables that are not caused by T , M and Y such that $0 < P(T = t|X) < 1$ and $0 < P(M(t) = m|T = t, X) < 1$ holds for all $x \in \text{supp}(X)$ and $m \in \text{supp}(M)$.

Under Sequential Ignorability [A-3](#), it is easy to show that the distributions of counterfactual variables are identified by $P(Y(t, m)|X) = P(Y|X, T = t, M = m)$ and $P(M(t)|X) = P(M|X, T = t)$ and thereby the mediating causal effects can be expressed as:

$$ADE(t) = \int \left(E(Y|T = t_1, M = m, X = x) - E(Y|T = t_0, M = m, X = x, X = x) \right) dF_{M|T=t, X=x}(m) dF_X(x) \quad (71)$$

$$AIE(t) = \int \left(E(Y|T = t, M = m, X = x) \left[dF_{M|T=t_1, X=x}(m) - dF_{M|T=t_0, X=x}(m) \right] \right) dF_X(x). \quad (72)$$

Imai, Tingley, Keele and Yamamoto offer a substantial line of research that explores the identifying properties of Sequential Ignorability Assumption [A-3](#). See [Imai, Keele, Tingley and Yamamoto \(2011a\)](#) for a comprehensive discussion of the benefits and limitations of the sequential ignorability assumption.

The main critics of Sequential Ignorability [A-3](#) is that it does not hold under the presence of either *Confounders* or *Unobserved Mediators* ([Heckman and Pinto, 2015](#)).

The independence relation (69) assumes that T is exogenous conditioned on X . There exists no unobserved variable that causes T and Y or T and M . For instance, the Sequential Ignorability [A-3](#) holds for the model defined in (??) because:

$$(\epsilon_Y, \epsilon_M) \perp\!\!\!\perp \epsilon_T \Rightarrow (f_Y(t', m, \epsilon_Y), f_M(t, \epsilon_M)) \perp\!\!\!\perp f_T(\epsilon_T) \Rightarrow (Y(t', m), M(t)) \perp\!\!\!\perp T. \quad (73)$$

$$\epsilon_Y \perp\!\!\!\perp \epsilon_M | \epsilon_T \Rightarrow f_Y(t', m, \epsilon_Y) \perp\!\!\!\perp f_M(t, \epsilon_M) | f_T(\epsilon_T) \Rightarrow Y(t', m) \perp\!\!\!\perp M(t) | T, \quad (74)$$

where the initial independence relation in (73) and (74) comes from the independence of error terms.

This assumption is expected to hold in experimental data when treatment T is randomly assigned. The independence relation (70) assumes that M is exogenous conditioned on X and T . It assumes that no confounding variable causing M and Y . Sequential Ignorability [A-3](#) is an extension of the Ignorability Assumption of [Rosenbaum and Rubin \(1983\)](#) that also assumes that a treatment T is exogenous when conditioned on pre-treatment variables. [Robins \(2003\)](#); [Petersen, Sinisi and Van der Laan \(2006\)](#); [Rubin \(2004\)](#) state similar identifying criteria that assume no confounding variables. Those assumptions are not testable.

Sequential Ignorability [A-3](#) assumes that: (1) the confounding variable V is observed, that is, the pre-treatment variables X ; and (2) that there is no unobserved mediator U . This assumption is unappealing for many because it solves the identification problem generated by confounding variables by assuming that those do not exist ([Heckman, 2008](#)).

Consider a change in the treatment variable T denoted by $\Delta(t) = t_1 - t_0$. The Direct and indirect effects can be expressed by:

$$\begin{aligned} ADE(t') &= \left(\lambda_{YT} \cdot t_1 + \lambda_{YM} \cdot E(M(t')) \right) - \left(\lambda_{YT} \cdot t_0 + \lambda_{YM} \cdot E(M(t')) \right) \\ \therefore ADE &= \lambda_{YT} \cdot \Delta(t) \end{aligned} \tag{75}$$

$$\begin{aligned} \text{and } AIE(t') &= \left(\lambda_{YT} \cdot t' + \lambda_{YM} \cdot E(M(t_1)) \right) - \left(\lambda_{YT} \cdot t' + \lambda_{YM} \cdot E(M(t_0)) \right) \\ &= \left(\lambda_{YT} \cdot t' + \lambda_{YM} \lambda_M \cdot t_1 \right) - \left(\lambda_{YT} \cdot t' + \lambda_{YM} \lambda_M \cdot t_0 \right) \\ \therefore AIE &= \lambda_{YM} \cdot \lambda_M \cdot \Delta(t) \end{aligned} \tag{76}$$

Online Appendix C The IV Mediate Model under Assumption A-2

Panel A of Table [Online Appendix Table 1](#) represents the causal relations of the mediation model (1)–(3) as a directed acyclic graph (DAG). Squares represent observed variables, while circles denote unobserved variables. Causal relations are denoted by solid lines while the dependence structure among error variables is depicted by dashed lines. Table [Online Appendix Table 1](#) is a version of *Model III* in Table 1 that accounts for the statistical dependence among error terms in [A-2](#).

Table Online Appendix Table 1: The Mediation Model with IV

A. DAG Representation

B. Model Equations

$$\text{Treatment variable: } T = f_T(Z, \epsilon_T)$$

$$\text{Observed mediator: } M = f_M(T, \epsilon_M)$$

$$\text{Outcome: } Y = f_Y(T, M, \epsilon_Y)$$

$$\text{where: } \epsilon_T \not\perp \epsilon_M, \epsilon_M \not\perp \epsilon_Y, \epsilon_T \not\perp \epsilon_Y | \epsilon_M$$

$$\text{and: } Z \perp\!\!\!\perp (\epsilon_T, \epsilon_M, \epsilon_Y), \epsilon_T \perp\!\!\!\perp \epsilon_Y$$

Online Appendix D Identification of Causal Parameters

When we additionally allow for an unobserved mediator U that is caused by T and causes both M and Y (see **Remark 2.3**), the linear mediation model we investigate can be fully described by the following equations:

$$\text{Instrumental Variable } Z = \epsilon_Z, \quad (77)$$

$$\text{Treatment } T = \xi_Z \cdot Z + \xi_V \cdot V_T + \epsilon_T, \quad (78)$$

$$\text{Unobserved Mediator } U = \zeta_T \cdot T + \epsilon_U, \quad (79)$$

$$\text{Observed Mediator } M = \varphi_T \cdot T + \varphi_U \cdot U + \delta_Y \cdot V_Y + \delta_T \cdot V_T + \epsilon_M, \quad (80)$$

$$\text{Outcome } Y = \beta_T \cdot T + \beta_M \cdot M + \beta_U \cdot U + \beta_V \cdot V_Y + \epsilon_Y, \quad (81)$$

$$\text{Exogenous Variables } Z, V_T, V_M, \epsilon_Z, \epsilon_T, \epsilon_U, \epsilon_M, \epsilon_Y \text{ are statistically independent variables,} \quad (82)$$

$$\text{Scalar Coefficients } \xi_Z, \xi_V, \zeta_T, \varphi_T, \varphi_U, \delta_Y, \delta_T, \beta_T, \beta_M, \beta_U, \beta_V \quad (83)$$

$$\text{Unobserved Variables } V_T, V_M, U, \epsilon_Z, \epsilon_T, \epsilon_U, \epsilon_M, \epsilon_Y. \quad (84)$$

We assume that all variables have mean zero. This assumption does not incur in less of generality, but simplify notation as intercepts can be suppressed.

We first eliminate the unobserved mediator U from Equations (80)–(81) by iterated substitution. Equations (81)–(81) are then expressed as:

$$M = (\varphi_T + \varphi_U \zeta_T) \cdot T + \varphi_U \cdot \epsilon_U + \delta_Y \cdot V_Y + \delta_T \cdot V_T + \epsilon_M, \quad (85)$$

$$Y = (\beta_T + \beta_U \zeta_T) \cdot T + \beta_M \cdot M + \beta_U \cdot \epsilon_U + \beta_V \cdot V_Y + \epsilon_Y. \quad (86)$$

We use the following transformation of parameters to save on notation:

$$\tilde{\varphi}_T = \varphi_T + \varphi_U \zeta_T, \quad (87)$$

$$\tilde{\beta}_T = \beta_T + \beta_U \zeta_T, \quad (88)$$

$$\tilde{U} = \epsilon_U. \quad (89)$$

We use equations (85)–(89) to simplify Model (77)–(81) into the following equations:

$$\text{Instrumental Variable } Z = \epsilon_Z, \quad (90)$$

$$\text{Treatment } T = \xi_Z \cdot Z + \xi_V \cdot V_T + \epsilon_T, \quad (91)$$

$$\text{Observed Mediator } M = \tilde{\varphi}_T \cdot T + \varphi_U \cdot \tilde{U} + \delta_Y \cdot V_Y + \delta_T \cdot V_T + \epsilon_M, \quad (92)$$

$$\text{Outcome } Y = \tilde{\beta}_T \cdot T + \beta_M \cdot M + \beta_U \cdot \tilde{U} + \beta_V \cdot V_Y + \epsilon_Y. \quad (93)$$

In this linear model, the counterfactual outcomes $M(t), Y(t), Y(m), Y(m, t)$ are given by:

$$M(t) = \tilde{\varphi}_T \cdot t + \varphi_U \cdot \tilde{U} + \delta_Y \cdot V_Y + \delta_T \cdot V_T + \epsilon_M, \quad (94)$$

$$Y(m) = \tilde{\beta}_T \cdot T + \beta_M \cdot m + \beta_U \cdot \tilde{U} + \beta_V \cdot V_Y + \epsilon_Y. \quad (95)$$

$$Y(t, m) = \tilde{\beta}_T \cdot t + \beta_M \cdot m + \beta_U \cdot \tilde{U} + \beta_V \cdot V_Y + \epsilon_Y. \quad (96)$$

$$\begin{aligned} Y(t) &= \tilde{\beta}_T \cdot t + \beta_M \cdot M(t) + \beta_U \cdot \tilde{U} + \beta_V \cdot V_Y + \epsilon_Y \\ &= (\tilde{\beta}_T + \beta_M \tilde{\varphi}_T) \cdot t + (\beta_U + \beta_M \varphi_U) \cdot \tilde{U} + (\beta_V + \beta_M \delta_Y) \cdot V_Y + \beta_M \delta_T \cdot V_T + \beta_M \cdot \epsilon_M + \epsilon_Y. \end{aligned} \quad (97)$$

We claim that the coefficients associated with unobserved variables $V_T, \tilde{U}, V_Y$ may only be identified up a linear transformation. Consider the coefficients δ_T, β_V that multiply the unobserved variable V_T in Equations (91) and (92) respectively. Suppose a linear transformation that multiplies V_T by a constant $\kappa \neq 0$. The model would remain the same if coefficients δ_T, β_V were divided by the same constant κ . This is a typical fact in the literature of linear factor models. We solve this non-identification problem by impose that each unobserved variable $V_T, \tilde{U}, V_Y$ has unit variance:

$$\text{var}(V_T) = \text{var}(\tilde{U}) = \text{var}(V_Y) = 1. \quad (98)$$

Assumption (98) is typically termed as *anchoring* of unobserved factors in the literature of factor analysis. This assumption does not incur in any loss of generality for the identification of direct, indirect or total causal effects of T (and M) on Y as expressed in the following section.

Online Appendix D.1 Defining Causal Parameters

The literature of mediation analysis term relevant causal parameters as:

- Total Effect of T on Y , that is, $\frac{dE(Y(t))}{dt}$.
- Direct Effect of T on Y , that is $\frac{\partial E(Y(t, m))}{\partial t}$.
- Effect of M on Y , that is, $\frac{dE(Y(m))}{dm}$.
- Effect of T on M , that is, $\frac{dE(M(t))}{dt}$.
- Indirect Effect of T on Y , that is $\frac{\partial E(Y(t, m))}{\partial m} \cdot \frac{dE(M(t))}{dt}$.

According to the counterfactual variables in (94)–(97), these causal effects are given by:

$$\text{Total Effect of } T \text{ on } Y : \frac{dE(Y(t))}{dt} = \tilde{\varphi}_T \cdot \beta_M + \tilde{\beta}_T. \quad (99)$$

$$\text{Direct Effect of } T \text{ on } Y : \frac{\partial E(Y(t, m))}{\partial t} = \tilde{\beta}_T. \quad (100)$$

$$\text{Effect of } M \text{ on } Y : \frac{dE(Y(m))}{dm} = \beta_M. \quad (101)$$

$$\text{Effect of } T \text{ on } M : \frac{dE(M(t))}{dt} = \tilde{\varphi}_T. \quad (102)$$

$$\text{Indirect Effect of } T \text{ on } Y : \frac{\partial E(Y(t, m))}{\partial m} \cdot \frac{dE(M(t))}{dt} = \beta_M \cdot \tilde{\varphi}_T. \quad (103)$$

Online Appendix D.2 Identifying Equations

Model (90)–(93) can be conveniently expressed in matrix notation. In Equation (104) we define $\mathbf{X} = [Z, T, M, Y]'$ as the vector of observed variables, $\mathbf{V} = [V_T, V_Y, \tilde{U}]'$ as the vector of unobserved confounding variables, and $\boldsymbol{\epsilon} = [\epsilon_Z, \epsilon_T, \epsilon_M, \epsilon_Y]'$ as the vector of exogenous error terms. According to (82), the random vectors $\mathbf{V}$ and $\boldsymbol{\epsilon}$ are independent, that is, $\mathbf{V} \perp\!\!\!\perp \boldsymbol{\epsilon}$. We use $\mathbf{K}$ in (104) for the matrix of parameters that multiply $\mathbf{X}$ and $\mathbf{A}$ for the matrix of parameters that multiply $\mathbf{V}$.

$$\mathbf{X} = \begin{pmatrix} Z \\ T \\ M \\ Y \end{pmatrix}, \quad \mathbf{V} = \begin{pmatrix} V_T \\ V_Y \\ \tilde{U} \end{pmatrix}, \quad \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_Z \\ \epsilon_T \\ \epsilon_M \\ \epsilon_Y \end{pmatrix}, \quad \mathbf{K} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ \xi_Z & 0 & 0 & 0 \\ 0 & \tilde{\varphi}_T & 0 & 0 \\ 0 & \tilde{\beta}_T & \beta_M & 0 \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} 0 & 0 & 0 \\ \xi_V & 0 & 0 \\ \delta_Y & \delta_Y & \varphi_U \\ 0 & \beta_V & \beta_U \end{bmatrix}. \quad (104)$$

Using the notation in (104), we can express the linear system (90)–(93) as following:

$$\underbrace{\begin{pmatrix} Z \\ T \\ M \\ Y \end{pmatrix}}_{\mathbf{X}} = \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 \\ \xi_Z & 0 & 0 & 0 \\ 0 & \tilde{\varphi}_T & 0 & 0 \\ 0 & \tilde{\beta}_T & \beta_M & 0 \end{bmatrix}}_{\mathbf{K}} \cdot \underbrace{\begin{pmatrix} Z \\ T \\ M \\ Y \end{pmatrix}}_{\mathbf{X}} + \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ \xi_V & 0 & 0 \\ \delta_T & \delta_Y & \varphi_U \\ 0 & \beta_V & \beta_U \end{bmatrix}}_{\mathbf{A}} \cdot \underbrace{\begin{pmatrix} V_T \\ V_Y \\ \tilde{U} \end{pmatrix}}_{\mathbf{V}} + \underbrace{\begin{pmatrix} \epsilon_Z \\ \epsilon_T \\ \epsilon_M \\ \epsilon_Y \end{pmatrix}}_{\boldsymbol{\epsilon}}, \quad (105)$$

$$\mathbf{X} = \mathbf{K} \cdot \mathbf{X} + \mathbf{A} \cdot \mathbf{V} + \boldsymbol{\epsilon}. \quad (106)$$

The coefficients in matrices $\mathbf{K}$, $\mathbf{A}$ are identified through the covariance matrices of observed variables. We use $\Sigma_{\mathbf{X}} = \text{cov}(\mathbf{X}, \mathbf{X})$ for the covariance matrix of observed variables $\mathbf{X}$, and $\Sigma_{\boldsymbol{\epsilon}} = \text{cov}(\boldsymbol{\epsilon}, \boldsymbol{\epsilon})$ for the vector of error terms $\boldsymbol{\epsilon}$. $\Sigma_{\boldsymbol{\epsilon}}$ is a diagonal matrix due to statistical independence of error terms. We also use $\Sigma_{\mathbf{V}} = \text{cov}(\mathbf{V}, \mathbf{V})$ for the covariance of unobserved variables $\mathbf{V}$. The unobserved variables in $\mathbf{V}$ are statistically independent and have unit variance (98), thus $\Sigma_{\mathbf{V}} = \mathbf{I}$ where $\mathbf{I}$ is the identity matrix. Moreover, $\mathbf{V} \perp\!\!\!\perp \boldsymbol{\epsilon}$ implies that $\text{cov}(\mathbf{V}, \boldsymbol{\epsilon}) = \mathbf{0}$, where $\mathbf{0}$ is a matrix of elements zero.

Equation (109) determines the relation between the covariance matrices of observed and unobserved variables:

$$\mathbf{X} = \mathbf{K} \cdot \mathbf{X} + \mathbf{A} \cdot \mathbf{V} + \boldsymbol{\epsilon} \Rightarrow (\mathbf{K} - \mathbf{I}) \mathbf{X} = \mathbf{A} \cdot \mathbf{V} + \boldsymbol{\epsilon}, \quad (107)$$

$$\Rightarrow (\mathbf{K} - \mathbf{I}) \Sigma_{\mathbf{X}} (\mathbf{K} - \mathbf{I})' = \mathbf{A} \Sigma_{\mathbf{V}} \mathbf{A}' + \Sigma_{\boldsymbol{\epsilon}}, \quad (108)$$

$$\Rightarrow (\mathbf{K} - \mathbf{I}) \Sigma_{\mathbf{X}} (\mathbf{K} - \mathbf{I})' = \mathbf{A} \mathbf{A}' + \Sigma_{\boldsymbol{\epsilon}}, \quad (109)$$

where the second equation is due to $\mathbf{V} \perp\!\!\!\perp \boldsymbol{\epsilon}$ and the third equations comes from $\Sigma_{\mathbf{V}} = \mathbf{I}$.

Equation (109) generates ten equalities. Four equalities are due to the diagonal of the covariance matrices $(\mathbf{K} - \mathbf{I}) \Sigma_{\mathbf{X}} (\mathbf{K} - \mathbf{I})'$ and $\mathbf{A} \mathbf{A}' + \Sigma_{\boldsymbol{\epsilon}}$ in (109). The remaining six equalities from the off-diagonal relation of the covariance matrices in (109).

The diagonal elements of $\Sigma_{\boldsymbol{\epsilon}}$ are the variances of the error terms $\epsilon_Z, \epsilon_T, \epsilon_M, \epsilon_Y$. Thereby each diagonal equation generated by (109) adds one unobserved term to the system of quadratic equations. The point-identification of the model coefficients arises from the six off-diagonal equations

generated by (109). Those are listed below:

$$\text{cov}(Z, T) - \text{cov}(Z, Z) \cdot \xi_Z = 0 \quad (110)$$

$$\text{cov}(Z, M) - \text{cov}(Z, T) \cdot \tilde{\varphi}_T = 0 \quad (111)$$

$$\text{cov}(Z, Y) - \text{cov}(Z, M) \cdot \beta_M - \text{cov}(Z, T) \cdot \tilde{\beta}_T = 0 \quad (112)$$

$$\text{cov}(T, Y) - \text{cov}(T, T) \cdot \tilde{\beta}_T - \text{cov}(T, M) \cdot \beta_M = 0 \quad (113)$$

$$\text{cov}(M, Y) - \text{cov}(T, M) \cdot \tilde{\beta}_T - \text{cov}(M, M) \cdot \beta_M = \beta_U \cdot \varphi_U + \beta_V \cdot \delta_Y \quad (114)$$

$$\text{cov}(T, M) - \text{cov}(T, T) \cdot \tilde{\varphi}_T = \delta_T \cdot \xi_V \quad (115)$$

Simple manipulation of Equations (110)–(115) generate the identification of the following parameters:

$$\xi_Z = \frac{\text{cov}(Z, T)}{\text{cov}(Z, Z)} \quad \text{from Eq.(110)} \quad (116)$$

$$\tilde{\varphi}_T = \frac{\text{cov}(Z, M)}{\text{cov}(Z, T)} \quad \text{from Eq.(111)} \quad (117)$$

$$\beta_M = \frac{\text{cov}(Z, T) \text{cov}(T, Y) - \text{cov}(T, T) \text{cov}(Z, Y)}{\text{cov}(T, M) \text{cov}(Z, T) - \text{cov}(T, T) \text{cov}(Z, M)} \quad \text{from Eqs.(112)–(113)} \quad (118)$$

$$\tilde{\beta}_T = \frac{\text{cov}(Z, M) \text{cov}(T, Y) - \text{cov}(Z, Y) \text{cov}(T, M)}{\text{cov}(T, T) \text{cov}(Z, M) - \text{cov}(Z, T) \text{cov}(T, M)} \quad \text{from Eqs.(112)–(113)} \quad (119)$$

$$\beta_U \cdot \varphi_U + \beta_V \cdot \delta_Y = \text{cov}(M, Y) - \text{cov}(M, M) \cdot \beta_M - \text{cov}(T, M) \cdot \tilde{\beta}_T \quad \text{from Eq.(114)} \quad (120)$$

$$\delta_T \cdot \xi_V = \frac{\text{cov}(T, M) \text{cov}(Z, M) - \text{cov}(T, T) \text{cov}(Z, Y)}{\text{cov}(Z, M)} \quad \text{from Eq.(115)} \quad (121)$$

Moreover, if we divide Equation (112) by $\text{cov}(Z, T)$ we obtain:

$$\frac{\text{cov}(Z, Y)}{\text{cov}(Z, T)} - \frac{\text{cov}(Z, M)}{\text{cov}(Z, T)} \cdot \beta_M - \frac{\text{cov}(Z, T)}{\text{cov}(Z, T)} \cdot \tilde{\beta}_T = 0 \quad (122)$$

$$\Rightarrow \frac{\text{cov}(Z, Y)}{\text{cov}(Z, T)} - \tilde{\varphi}_T \cdot \beta_M - \tilde{\beta}_T = 0 \quad (123)$$

$$\Rightarrow \tilde{\varphi}_T \cdot \beta_M + \tilde{\beta}_T = \frac{\text{cov}(Z, Y)}{\text{cov}(Z, T)}. \quad (124)$$

The four causal of interest parameters defined in (99)–(102) are respectively identified by Equations (117), (118), (119) and (124):

$$\frac{dE(M(t))}{dt} = \tilde{\varphi}_T = \frac{\text{cov}(Z, M)}{\text{cov}(Z, T)}, \quad (125)$$

$$\frac{dE(Y(m))}{dm} = \beta_M = \frac{\text{cov}(Z, Y) \text{cov}(T, T) - \text{cov}(Y, T) \text{cov}(Z, T)}{\text{cov}(Z, M) \text{cov}(T, T) - \text{cov}(M, T) \text{cov}(Z, T)}, \quad (126)$$

$$\frac{\partial E(Y(t, m))}{\partial t} = \tilde{\beta}_T = \frac{\text{cov}(Z, M) \text{cov}(T, Y) - \text{cov}(Z, Y) \text{cov}(T, M)}{\text{cov}(T, T) \text{cov}(Z, M) - \text{cov}(Z, T) \text{cov}(T, M)}, \quad (127)$$

$$\frac{dE(Y(t))}{dt} = \tilde{\varphi}_T \cdot \beta_M + \tilde{\beta}_T = \frac{\text{cov}(Z, Y)}{\text{cov}(Z, T)}. \quad (128)$$

Next section explains that each causal effect (125)–(128) can be evaluated by standard Two-stage Least Squares regressions.

Online Appendix E Exploring Alternative Approaches

We investigate the mediation model in which the treatment variable T and the mediator variable M are endogenous. Our solution imposes causal relations among unobserved variables that enable the identification of three causal effects using only one dedicated instrument for T .

Our method contrasts to two broad alternative approaches to gaining identification in mediation analysis. One of these is to assume that the treatment T and the mediator M are exogenous given observed variables (Imai et al., 2010, 2011a,b).⁶⁵ In this case, treatment T is as good as randomly assigned and the resulting model is equivalent to assuming no confounding variables and no unobserved mediators U in *Model III* of Table 1.⁶⁶ Relatedly, Yamamoto (2014) studies the case of a binary treatment indicator and a single instrument, assuming that the instrument Z is independent of the counterfactual outcome $Y(m, t)$ and that the mediator variables is exogenous conditioned on treatment compliance.⁶⁷

A second class of models relies on additional instrumental variables dedicated to the mediator M . Powdthavee (2009); Burgess, Daniel, Butterworth and Thompson (2015) and Jhun (2015) achieve identification using two instruments and parametric assumptions that shape the endogeneity of T and M . Two important contributions to this literature that use non-parametric identification are Frölich and Huber (2017) and Jun et al. (2016).⁶⁸ This second class of models does not assume away confounding effects; i.e. variables T, M, Y remain endogenous. It thus constitutes an alternative approach to our identification problem, which is to seek for another instrument that is dedicated to M .⁶⁹ Because of its natural appeal, we discuss this approach here and contrast its identification requirements explicitly to ours. A standard mediation model with confounding variables V and two separate dedicated instrumental variables (for separate endogenous variables) is described as follows:

$$\text{Treatment variable: } T = f_T(Z_T, V, \epsilon_T), \quad (129)$$

$$\text{Observed Mediator: } M = f_M(T, Z_M, V, \epsilon_M), \quad (130)$$

$$\text{Outcome: } Y = f_Y(T, M, V, \epsilon_Y), \quad (131)$$

$$\text{where: } (Z_T, Z_M) \perp\!\!\!\perp V. \quad (132)$$

This model is presented as a DAG in Table Online Appendix Table 2. In this model, the exclusion restriction $Z_M \perp\!\!\!\perp Y(m)$ and also $Z_M \perp\!\!\!\perp Y(m)|T$ hold. Thereby Z_M can be used to evaluate the causal effects of M on Y .⁷⁰

The empirical challenge in evaluation Model (129)–(132) is to find a suitable candidate for Z_M . There are three potential concerns with any dedicated instrument for M : (i) Z_M may correlate with V ; (ii) Z_M may directly affect Y ; and (iii) Z_M may correlate with Z_T . Concerns (i) and (ii) define the usual requirements for any valid instrument to identify the effect of M on Y . The latter concern (iii)

⁶⁵Robins and Greenland (1992) and Geneletti (2007) consider instruments that perfectly correlate with the mediator variable such that the exogeneity condition still holds.

⁶⁶If the treatment T were indeed randomly assigned, then one could use the interaction of the treatment with observed covariates as instruments to identify the causal effect of M on Y . Versions of this approach are examined in Ten Have, Joffe, Lynch, Brown, Maisto and Beck (2007); Dunn and Bentall (2007); Small (2012); Gennetian, Bos and Morris (2002).

⁶⁷In our notation, this means that $Y(m, t) \perp\!\!\!\perp Z$ and $Y(t, m) \perp\!\!\!\perp M(t)|(T, P = c)$, where T denotes treatment assignment and P stands for an indicator of treatment compliance. Neither assumption holds in *Model III* or *Model IV* of Table 1.

⁶⁸Both papers examine the effect of a binary indicator for treatment T . Frölich and Huber (2017) relies on two dedicated instruments (for T and M) and a monotonicity restriction with respect to M . Jun et al. (2016) uses three dedicated instruments but does not require the monotonicity restriction.

⁶⁹Recently, Frölich and Huber (2017) provide an important contribution on the mediation model with two dedicated instruments.

⁷⁰If T were to cause Z_M , then only $Z_M \perp\!\!\!\perp Y(m)|T$ would hold.

Table Online Appendix Table 2: General Mediation Model and Violation of Exclusion Restrictions

A. Directed Acyclic Graph (DAG) Representation

General IV Model with Two Instruments Violations of the Exclusion Restriction

The left figure gives the directed acyclic graph (DAG) representation of the general IV Model with two dedicated instruments. The right figure gives the same DAG, but also depicts the identification concerns discussed in the body of the text.

is specific to the mediation context. The three concerns are depicted as dashed errors in the right figure of Table [Online Appendix Table 2](#). A potential candidate for Z_M is automation. Automation, i.e. replacing workers with machines, robots and computer-assisted technologies, is usually viewed as the ‘other big shock’ that has hit high-wage labor markets in the last decades. For example, [Acemoglu and Restrepo \(2017\)](#) estimate that an additional robot per thousand workers has reduced employment in the U.S. by about 0.18–0.34 percentage points and wages by 0.25–0.50 percent. The effects of automation are not expected to abate. For example, [Frey and Osborne \(2017\)](#) predict that 47% of U.S. workers are at risk of automation over the next two decades. In brief, automation has had and will likely continue to have substantial effects on labor market outcomes M and therefore seems like a good candidate dedicated instrument Z_M . We view concern (iii) as addressed in this context because [Autor et al. \(2015\)](#) have provided convincing evidence that automation and import exposure are largely orthogonal, making the two forces separable in the data at both the industry-level and the regional level. Concern (i) still is that firms may automate in response to other unobserved factors that could directly impact their labor demand. Indeed, firm-level technology upgrading does appear to respond to the China shock as shown by [Bloom, Draca and Van Reenen \(2016\)](#). This violates the independence $Z_M \perp\!\!\!\perp V$ in (132) and thereby the exclusion restriction $Z_M \perp\!\!\!\perp Y(m)|T$ does not hold. However, this concern may again be largely addressed if we think of Z_M not as actually measured technology upgrading but as some more exogenous measure, e.g. *exposure to robot adoption* as in [Acemoglu and Restrepo \(2017\)](#) or employment-weighted occupational measures like *routine task intensity* ([Autor and Dorn, 2013](#)) or *automatability* ([Frey and Osborne, 2017](#)). In our empirical context, concern (ii)—automation could impact voting behavior through channels other than M —is the most worrisome, and in fact clearly disqualifies automation as a dedicated instrument for M . While a German assembly-line worker will likely neither observe nor care about Australian imports of Chinese consumer electronics (i.e. Z_T), he/she will not only be aware of the potential automatability of their assembly-line job (i.e. Z_M) but may indeed seek out a more protectionist political agenda in anticipation of automation’s consequences, i.e. even before any detrimental effects in the labor market.

Online Appendix F Estimation of Causal Parameters

Our goal is to show that the four causal parameters listed in Equations (125)–(128) can be estimated using the standard Two-stage Least Square (2SLS) estimator. We revise the standard equations of the 2SLS estimators for sake of completeness.

Equations (133)–(134) present the first and stages of a generic 2SLS regression in which T stands for the endogenous variable, Z is the instrumental variable and Y is the targeted outcome.

$$\text{First Stage: } T = \kappa_1 + \beta_1 \cdot Z + \epsilon_1, \quad (133)$$

$$\text{Second Stage: } Y = \kappa_2 + \beta_2 \cdot T + \epsilon_2. \quad (134)$$

The 2SLS estimator relies on the assumptions that the instrument Z is statistically independent of the term ϵ_2 while T is not. It is well-known that the 2SLS estimator $\hat{\beta}_2$ is given by the ratio of the sample covariances $\text{cov}(Z, Y)$ and $\text{cov}(Z, T)$. Moreover $\hat{\beta}_2$ is a consistent estimator of parameter β_2 :

$$\text{plim}(\hat{\beta}_2) = \frac{\text{cov}(Z, Y)}{\text{cov}(Z, T)} = \beta_2. \quad (135)$$

Consider the inclusion of additional covariates X in both stages of the 2SLS method. Variables X in (136)–(137) play the role of control covariates in the first stage and second stages of the 2SLS estimator. Control covariates X directly causes Y in (137) while the instrument Z only causes Y though it impact on T .

$$\text{First Stage: } T = \kappa_1 + \beta_1 \cdot Z + \psi_1 \cdot X + \epsilon_1, \quad (136)$$

$$\text{Second Stage: } Y = \kappa_2 + \beta_2 \cdot T + \psi_2 \cdot X + \epsilon_2. \quad (137)$$

The 2SLS model (136)–(137) relies on the assumption that the instrument Z and control covariates X are independent of error term ϵ_2 , that is, $(Z, X) \perp\!\!\!\perp \epsilon_2$. The 2SLS estimator $\hat{\beta}_2$ for parameter β_2 is expressed by Equation (138) and it is a consistent estimator under model assumptions.

$$\text{plim}(\hat{\beta}_2) = \frac{\text{cov}(Z, Y) \text{cov}(X, X) - \text{cov}(Y, X) \text{cov}(Z, X)}{\text{cov}(Z, T) \text{cov}(X, X) - \text{cov}(T, X) \text{cov}(Z, X)} = \beta_2. \quad (138)$$

The 2SLS estimator $\hat{\psi}_2$ for parameter ψ_2 is expressed by Equation (139) and it is a consistent estimator under model assumptions.

$$\text{plim}(\hat{\psi}_2) = -\frac{\text{cov}(Z, Y) \text{cov}(T, X) - \text{cov}(Y, X) \text{cov}(Z, T)}{\text{cov}(Z, T) \text{cov}(X, X) - \text{cov}(T, X) \text{cov}(Z, X)} = \psi_2. \quad (139)$$

Each of the identification formulas for the causal effects in (125)–(128) describes a ratio of covariances that corresponds to one of the three 2SLS formulas (135), (138) or (139).

The effect of choice T on mediator M is given by:

$$\frac{dE(M(t))}{dt} = \tilde{\varphi}_T = \frac{\text{cov}(Z, M)}{\text{cov}(Z, T)}.$$

According to Equation (135), this effect can be estimated by the 2SLS regression (133)–(134) in which Z is the instrument, T is the endogenous variable and M is the outcome.

The total effect of T on outcome Y is given by:

$$\frac{dE(Y(t))}{dt} = \tilde{\varphi}_T \cdot \beta_M + \tilde{\beta}_T = \frac{\text{cov}(Z, Y)}{\text{cov}(Z, T)}.$$

According to Equation (135), this effect can be estimated by the 2SLS regression (133)–(134) in which Z is the instrument, T is the endogenous variable and Y is the outcome.

The causal effect of mediator M on outcome Y is given by:

$$\frac{dE(Y(m))}{dm} = \beta_M = \frac{\text{cov}(Z, Y) \text{cov}(T, T) - \text{cov}(Y, T) \text{cov}(Z, T)}{\text{cov}(Z, M) \text{cov}(T, T) - \text{cov}(M, T) \text{cov}(Z, T)},$$

which can be estimated by the 2SLS regression (133)–(134) where Z is the instrument, T is the endogenous variable and M is the outcome.

The causal effect of mediator M on outcome M is given by:

$$\frac{dE(Y(m))}{dm} = \beta_M = \frac{\text{cov}(Z, Y) \text{cov}(T, T) - \text{cov}(Y, T) \text{cov}(Z, T)}{\text{cov}(Z, M) \text{cov}(T, T) - \text{cov}(M, T) \text{cov}(Z, T)}.$$

According to the 2SLS estimator in (138), this causal effect can be estimated by $\hat{\beta}_2$ in the 2SLS regression (136)–(137) in which Z plays the role of the instrument, M is the endogenous variable, T is the control covariate and Y is the outcome.

The Indirect Effect of choice T on outcome Y is given by:

$$\frac{\partial E(Y(t, m))}{\partial m} = \tilde{\beta}_T = \frac{\text{cov}(Z, M) \text{cov}(T, Y) - \text{cov}(Z, Y) \text{cov}(T, M)}{\text{cov}(T, T) \text{cov}(Z, M) - \text{cov}(Z, T) \text{cov}(T, M)}.$$

According to the 2SLS estimator in (139), this causal effect can be estimated by $\hat{\psi}_2$ in the 2SLS regression (136)–(137) in which Z plays the role of the instrument, M is the endogenous variable, T is the control covariate and Y is the outcome.

Online Appendix G Total, Indirect and Direct Effects under One Instrument

Online Appendix D.2 describes a linear mediation model whose primary causal effects are identified by the following equations:

$$\text{Total Effect of } T \text{ on } Y : \frac{dE(Y(t))}{dt} = \frac{\text{cov}(Z, Y)}{\text{cov}(Z, T)}. \quad (140)$$

$$\text{Direct Effect of } T \text{ on } Y : \frac{\partial E(Y(t, m))}{\partial t} = \frac{\text{cov}(Z, M) \text{cov}(T, Y) - \text{cov}(Z, Y) \text{cov}(T, M)}{\text{cov}(T, T) \text{cov}(Z, M) - \text{cov}(Z, T) \text{cov}(T, M)}. \quad (141)$$

$$\text{Effect of } M \text{ on } Y : \frac{\partial E(Y(t, m))}{\partial m} = \frac{\text{cov}(Z, T) \text{cov}(T, Y) - \text{cov}(T, T) \text{cov}(Z, Y)}{\text{cov}(T, M) \text{cov}(Z, T) - \text{cov}(T, T) \text{cov}(Z, M)}. \quad (142)$$

$$\text{Effect of } T \text{ on } M : \frac{dE(M(t))}{dt} = \frac{\text{cov}(Z, M)}{\text{cov}(Z, T)}. \quad (143)$$

$$\text{Indirect Effect of } T \text{ on } Y : \frac{\partial E(Y(t, m))}{\partial m} \cdot \frac{dE(M(t))}{dt}. \quad (144)$$

The literature of mediation analysis typically expresses the total effect of T on Y as the sum of its direct and indirect effects. In our notation, this decomposition is stated as following:

$$\underbrace{\frac{dE(Y(t))}{dt}}_{\text{Total Effect}} = \underbrace{\frac{\partial E(Y(t, m))}{\partial t}}_{\text{Direct Effect}} + \underbrace{\frac{\partial E(Y(t, m))}{\partial m} \cdot \frac{dE(M(t))}{dt}}_{\text{Indirect Effect}}. \quad (145)$$

We show that the decomposition described in (145) is exact in the case of a single instrument. That is to say that the covariance ratio that identifies the total effect of T on Y in equation (140) is equal to the covariance ratio that identifies the direct effect in Equations (141) plus the multiplication of the covariance ratios that identify the effect of T on M in (143) and the effect of M on Y

described in Equation (142). We thank David Slichter for pointing out this fact.

$$\begin{aligned}
& \underbrace{\frac{\partial E(Y(t, m))}{\partial t}}_{\text{Direct Effect}} + \underbrace{\frac{\partial E(Y(t, m))}{\partial m} \cdot \frac{dE(M(t))}{dt}}_{\text{Indirect Effect}} \\
&= \underbrace{\frac{\text{cov}(Z, M) \text{cov}(T, Y) - \text{cov}(Z, Y) \text{cov}(T, M)}{\text{cov}(T, T) \text{cov}(Z, M) - \text{cov}(Z, T) \text{cov}(T, M)}}_{\frac{\partial E(Y(t, m))}{\partial t}} + \underbrace{\frac{\text{cov}(Z, T) \text{cov}(T, Y) - \text{cov}(T, T) \text{cov}(Z, Y)}{\text{cov}(T, M) \text{cov}(Z, T) - \text{cov}(T, T) \text{cov}(Z, M)}}_{\frac{\partial E(Y(t, m))}{\partial m}} \cdot \underbrace{\frac{\text{cov}(Z, M)}{\text{cov}(Z, T)}}_{\frac{dE(M(t))}{dt}} \\
&= \frac{\text{cov}(Z, M) \text{cov}(T, Y) - \text{cov}(Z, Y) \text{cov}(T, M)}{\text{cov}(T, T) \text{cov}(Z, M) - \text{cov}(Z, T) \text{cov}(T, M)} + \frac{\text{cov}(Z, M) \text{cov}(T, Y) - \text{cov}(T, T) \text{cov}(Z, Y)}{\text{cov}(T, M) \text{cov}(Z, T) - \text{cov}(T, T) \text{cov}(Z, M)} \frac{\text{cov}(Z, M)}{\text{cov}(Z, T)} \\
&= \frac{\text{cov}(Z, M) \text{cov}(T, Y) - \text{cov}(Z, Y) \text{cov}(T, M)}{\text{cov}(T, T) \text{cov}(Z, M) - \text{cov}(Z, T) \text{cov}(T, M)} + \frac{\text{cov}(T, T) \text{cov}(Z, Y) \frac{\text{cov}(Z, M)}{\text{cov}(Z, T)} - \text{cov}(Z, M) \text{cov}(T, Y)}{\text{cov}(T, T) \text{cov}(Z, M) - \text{cov}(Z, T) \text{cov}(T, M)} \\
&= \frac{\text{cov}(T, T) \text{cov}(Z, Y) \frac{\text{cov}(Z, M)}{\text{cov}(Z, T)} - \text{cov}(Z, Y) \text{cov}(T, M)}{\text{cov}(T, T) \text{cov}(Z, M) - \text{cov}(Z, T) \text{cov}(T, M)} \\
&= \frac{\text{cov}(T, T) \text{cov}(Z, M) \frac{\text{cov}(Z, Y)}{\text{cov}(Z, T)} - \text{cov}(Z, Y) \text{cov}(T, M)}{\text{cov}(T, T) \text{cov}(Z, M) - \text{cov}(Z, T) \text{cov}(T, M)} \\
&= \frac{\text{cov}(T, T) \text{cov}(Z, M) \frac{\text{cov}(Z, Y)}{\text{cov}(Z, T)} - \text{cov}(Z, Y) \text{cov}(T, M) \frac{\text{cov}(Z, T)}{\text{cov}(Z, T)}}{\text{cov}(T, T) \text{cov}(Z, M) - \text{cov}(Z, T) \text{cov}(T, M)} \\
&= \frac{\text{cov}(T, T) \text{cov}(Z, M) \frac{\text{cov}(Z, Y)}{\text{cov}(Z, T)} - \text{cov}(Z, T) \text{cov}(T, M) \frac{\text{cov}(Z, Y)}{\text{cov}(Z, T)}}{\text{cov}(T, T) \text{cov}(Z, M) - \text{cov}(Z, T) \text{cov}(T, M)} \\
&= \left(\frac{\text{cov}(T, T) \text{cov}(Z, M) - \text{cov}(Z, T) \text{cov}(T, M)}{\text{cov}(T, T) \text{cov}(Z, M) - \text{cov}(Z, T) \text{cov}(T, M)} \right) \cdot \left(\frac{\text{cov}(Z, Y)}{\text{cov}(Z, T)} \right) \\
&= \frac{\text{cov}(Z, Y)}{\text{cov}(Z, T)} = \underbrace{\frac{dE(Y(t))}{dt}}_{\text{Total Effect}}.
\end{aligned}$$

The first equality expresses the total effect of T on Y in terms of its direct and indirect effects. The second equality substitutes the direct and indirect effects by their identification formulas described in (141), (142) and (140). The third equation isolates and eliminates the common term $\text{cov}(Z, M)$ in the denominator of $\frac{dE(Y(m))}{dm}$. The fourth equation flips the sign of the terms in the last covariance ratio. Now the overall sum has the same denominator. The fifth equation eliminates the common term in the sum of the numerators of both ratios. The sixth equation exchange the covariances $\text{cov}(Z, M)$ and $\text{cov}(Z, Y)$ of the first term of the numerator. The seventh equation includes the term $\frac{\text{cov}(Z, T)}{\text{cov}(Z, T)}$ which is equal to one. The eighth equation exchange the covariances $\text{cov}(Z, Y)$ and $\text{cov}(Z, T)$ of the second term of the numerator. The ninth equation isolates the common denominator of the expression. The tenth equation eliminates the common first term of both numerator and denominator. The resulting formula is the covariate ratio $\frac{\text{cov}(Z, Y)}{\text{cov}(Z, T)}$ which, according to (140), is equal to the total effect of choice T on outcome Y .

Online Appendix H Controlling for Additional Covariates

The identification results and estimation methods of Section 2.3 can be easily extended to enable conditioning on additional covariates. Let $\mathbf{K}$ denotes a set of variables that we wish to control for. By this, we mean that variables $\mathbf{K}$ cause observed variables $\mathbf{X} = [Z, T, M, Y]'$. In other words, part of the covariation in $\mathbf{X}$ is not due to the causal relations of model (12)–(15), but rather due to the correlation induced by covariates $\mathbf{K}$. Thus, our task summates to isolate the portion of covariance $\Sigma_{\mathbf{X}}$ that is caused by variables $\mathbf{K}$.

A common approach is to use orthogonal projections to decompose a covariance matrix $\Sigma_{\mathbf{X}}$ into the part due to variables $\mathbf{K}$ and its complement. Let $\mathcal{K}$ be the linear space spanned by variables $\mathbf{K}$, and let $\mathcal{K}^\perp$ be the orthogonal complement to $\mathcal{K}$. Thus, the covariance $\Sigma_{\mathbf{X}}$ can be decomposed as follows:

$$\Sigma_{\mathbf{X}} = \Sigma_{\mathbf{X}|\mathcal{K}} + \Sigma_{\mathbf{X}|\mathcal{K}^\perp} \quad (146)$$

$$\text{where } \Sigma_{\mathbf{X}|\mathcal{K}^\perp} = \Sigma_{\mathbf{X}} - \Sigma_{\mathbf{X},\mathbf{K}} \Sigma_{\mathbf{K}}^{-1} \Sigma'_{\mathbf{X},\mathbf{K}}. \quad (147)$$

We can control for covariates $\mathbf{K}$ by simply replacing the covariance matrix $\Sigma_{\mathbf{X}}$ in (18) by $\Sigma_{\mathbf{X}|\mathcal{K}^\perp}$ in (147). All previous identification results follow. In practice, replacing $\Sigma_{\mathbf{X}}$ by $\Sigma_{\mathbf{X}|\mathcal{K}^\perp}$ is equivalent to adding covariates $\mathbf{K}$ as conditioning variables in the linear regressions (27)–(31) that estimate the model coefficients. For instance, instead of (27), parameter β_T^Z can now be estimated by the following OLS regression:

$$\text{New OLS for } \beta_T^Z: \quad T = \beta_T^Z \cdot Z + \beta_T^K \cdot K + \epsilon_T, \quad (148)$$

and equivalently for regressions (28)–(31).⁷¹ Controlling for additional covariates can be also understood as application of the Frisch-Waugh-Lovell Theorem where each of our main variables Z, T, M, Y are replaced by the residual generated by regressing these variables on covariates $\mathbf{K}$.

⁷¹ Equations (37) and (38) in Section 4 are the empirical equivalent of equations (28) and (29), with controls K added. Equations (39) and (40) in Section 5 are the empirical equivalent of equations (30) and (31), again with controls K added.

Online Appendix I Data Sources

Online Appendix I.1 Election Data

We focus on federal elections (*Bundestagswahlen*) because the timing of state elections (*Landtagswahlen*) and local elections (*Kommunalwahlen*) varies widely across German regions. Federal elections took place in 1987, in December 1990 after the reunification on October 3, and in 1994, 1998, 2002, 2005, and 2009. We define the first-period outcomes as changes in the vote-share from 1987 to 1998, and second-period outcomes as changes from 1998 to 2009. Election outcomes are observed at the level of 322 West German districts (*Landkreis*) in Period 1 and Period 2.

The average vote share of extreme-right parties is persistently below 5 percent in both periods. This presented a major challenge for our data collection, since official election statistics do not report all votes shares below the 5 percent minimum threshold separately by party. To extract information on extreme-right parties from this residual category, we had to contact the statistical offices of the German states and digitize some results from hard copies. By doing so, we have generated a unique data set that provides detailed insight into Germany's political constellation and allows us to create a precise measure of spatial variation in preferences also for fringe parties. This measure eventually allows us to extend existing studies on spatial variation of extreme-right activities and partisanship that were typically bound to the state level (Falk, Kuhn and Zweimüller (2011), Lubbers and Scheepers (2001)) or limited in their time horizon (Krueger and Pischke (1997)) to a new level of detail.

Online Appendix I.2 Trade Data

Our trade data stem from the U.N. Commodity Trade Statistics Database (Comtrade). The database provides information on trade flows between country pairs, detailed by commodity type. As in Dauth et al. (2014), we express all trade flows in thousands and convert them to 2005 Euros. To merge four-digit SITC2 product codes with our three-digit industry codes, we use a crosswalk provided by Dauth et al. (2014), who themselves employ a crosswalk provided by the U.N. Statistics Division to link product categories to NACE industries. In 92 percent of the cases, commodities map unambiguously into industries. For ambiguous cases, we use national employment shares from 1978 to partition them to industries. In this way, we end up with 157 manufacturing industries (excluding fuel products), classified according to the WZ73 industry classification.

Online Appendix I.3 Labor Market Data

We obtain information on local labor markets from two different sources. Information on employment, education, and the share of foreigners stems from the Social Security records in Germany.⁷² Based on the Social Security records, we calculate the import exposure measures for local labor markets, the share of high-skilled workers (with a tertiary degree), foreign workers, workers in the automobile or chemical industry, and wages. For the years before 1999, social security data are recorded at the place of work only. After 1999, place-of-work and place-of-residence information is available.

The remaining variables are provided by the German Federal Statistical Office. These variables include the overall population, the female population share, the population share of individuals

⁷²See Bender et al. (2000) for a detailed description of the data from the Institute for Employment Research (IAB). For an additional description of the regional distribution of wages across German municipalities, see Falck, Heblich and Otto (2013)

of working age (aged 18 to 65), the population share of individuals older than 65, and the unemployment rate, which is calculated by dividing the number of unemployed individuals by the working-age population.

Online Appendix J Background on German Politics 1987 to 2009

Online Appendix J.1 The German Election System

Since the end of WWII, Germany has had a multiparty system, with the two largest parties—the *Christian Democratic Union* (CDU) and the *Social Democratic Party of Germany* (SPD)—forming coalitions with either the *Free Democratic Party* (FDP) or the Greens (*Bündnis 90/Die Grünen*) during our observation period (1987 to 2009).⁷³ German elections are based on the principle of proportionality. The main vote, called the “second vote” (*Zweitstimme*), is being cast for parties but not for individual candidates.⁷⁴ We will exclusively focus on this party vote. The overall number of parliamentary seats is determined in proportion to a party’s share of the second vote. Parties further have to surpass a 5 percent minimum threshold to be represented in federal parliament. However, this does not mean that small parties do not capture any votes. Small parties that failed to pass the 5 percent threshold still captured about 11 percent of the total votes in our election data.

Online Appendix J.2 The Political Party Spectrum in Germany

We always classify the CDU, the SPD, the FDP, and the Greens as established parties. The conservative CDU and the social-democratic SPD are the dominant parties in Germany, in terms of both membership and votes obtained. For our period of analysis, one of those two parties was always in power. The liberal FDP participated in governments led by the CDU. The Greens are, for ideological reasons, usually the SPD’s preferred coalition partner. On the extreme right of the political spectrum, three parties have regularly run in federal elections. The National Democratic Party of Germany (NPD - *Nationaldemokratische Partei Deutschlands*), founded in 1964, the Republicans (REP - *Die Republikaner*), founded in 1983, and the German People’s Union (DVU - *Deutsche Volksunion*), founded in 1987 (and merged with the NPD in 2011).⁷⁵ They all follow neo-Nazi ideologies, are anti-democratic, polemicize against globalization, and agitate against immigrants and foreigners. All three have been monitored by the German Federal Office for the Protection of the Constitution (*Verfassungsschutz*). None of these extreme-right parties has ever passed the 5 percent hurdle required to enter Germany’s national parliament, and it is unthinkable that any mainstream party would ever form a coalition with them (see Art (2007)). On the far left of the political spectrum, there are around 10 parties and factions that are often related with each other. Besides the left party (*Die Linke*) and its predecessors, the *Party of Democratic Socialism* (PDS) and *Labour and Social Justice The Electoral Alternative* (WASG), three branches have been dominant: Successors to the Communist Party of Germany, which had been outlawed in 1956, e.g., the *German Communist Party* (DKP) and the *Communist Party of Germany* (KPD); Leninist, Stalinist, and Maoist organizations like the *Marxist-Leninist Party of Germany* (MLPD); and Trotskyist organizations

⁷³In this paper, we will always report votes for the CDU and its Bavarian subsection *Christian Social Union* (CSU) as combined CDU votes and refer to it as the CDU.

⁷⁴Voters can additionally elect individual candidates on a first-past-the-post basis. Ironically, this second ballot is called the “primary vote” (*Erststimme*). In every election district, the candidate who wins the majority of primary votes is directly elected to parliament. However, electoral law ensures that this has no significant effect on the overall distribution of seats, which is determined by the second vote.

⁷⁵In [Online Appendix J.4](#), we provide a history of these three parties. See also comprehensive work by Stöss (2010) or Mudde (2000).

such as the Party for Social Justice (PSG). Like the parties on the extreme right, these far-left parties are regularly monitored by either the Federal Office for the Protection of the Constitution or its state-level equivalents. We classify other parties that ran for elections but do not fit the above categories as small parties (see [Falck et al. 2014](#)).

Online Appendix J.3 Stance on Trade and Globalization

Both the large parties CDU and SPD have market-liberal as well as protectionist factions. In comparison, the CDU tends to be more market-friendly. Still, it was a government led by the SPD that implemented substantial labor market reforms in 2003-2005, amongst others decreasing employment protection, unemployment benefits, and establishing a low wage sector in Germany. The smaller FDP explicitly follows a market-liberal agenda, while the Green party focusses on environmental issues. More generally, the political left has traditionally been seen as opposing globalization and capturing the anti-globalization vote.⁷⁶ However, this is no unambiguous relationship, as the *The Economist* ([July 30th, 2016](#)) observes when headlining “Farewell, left versus right. The contest that matters now is open against closed.” Throughout Europe, the political left has found it difficult to take a coherent position against globalization in the last two decades, often hampered by internal intellectual conflicts ([Sommer 2008](#), [Arzheimer 2009](#)). In contrast, the right and far right successfully cobbled together an anti-globalization agenda ([Mughan et al., 2003](#)). For the case of Germany, [Sommer \(2008, p. 312\)](#) argues that “in opposing globalization, the left-wing usually criticizes an unjust and profit-oriented economic world order. [It] does not reject globalization per se but rather espouses a different sort of globalization. In contrast, the solutions proposed by the extreme right keep strictly to a national framework. The extreme right’s claim, therefore, that it is the only political force that opposes globalization fundamentally [...] rings true.” The following excerpt from the extreme-right NPD’s ‘candidate manual’ illustrates how Germany’s far right rolls protectionist anti-globalization themes into its broader nationalistic, anti-Semitic agenda: “Globalization is a planetary spread of the capitalist economic system under the leadership of the Great Money. This has, despite by its very nature being Jewish-nomadic and homeless, its politically and military protected location mainly on the East Coast of the United States” ([Grumke, 2012, p. 328](#)).⁷⁷

Online Appendix J.4 The Extreme Right in Germany

There is a strong sense of historical cultural roots and their time-persistence when it comes to explaining votes for far-right parties in Germany today. [Mocan and Raschke \(2014\)](#) use state-level survey aggregates from the ALLBUS, a general population survey for Germany, to show that people who live in states that had provided above-median support of the Nazi party in the 1928 elections have stronger anti-Semitic feelings today. [Voigtländer and Voth \(2015\)](#) use the same data to show that the effects of historical anti-Semitic attitudes on today’s political attitudes was amplified for the cohorts that grew up during Nazi Germany’s indoctrination programs in 1933–1945.

Having said that, there is substantial time-variation in the popularity of the far-right in Germany. The NPD, the oldest of the three major right-wing parties we consider, was founded in 1964 as the successor to the German Reich Party (DRP). Its goal was to unite a number of fragmented

⁷⁶To some extent this may still be the case. [Che et al. \(2016\)](#) for example argue that trade liberalization with China has turned American voters towards the Democrats, though it seems as if this might have not been true for the 2016 presidential elections.

⁷⁷The bundling of protectionist anti-globalization themes with xenophobic content has also been noted in the 2016 U.S. presidential election, see for example *The Guardian* ([March 7th 2016](#)).

far-right parties under one umbrella. Between 1966 and 1968, the NPD was elected into seven state parliaments, and in the 1969 federal election it missed the 5 percent minimum threshold by just 0.7 percentage points. Afterwards, support for the NPD declined and it took the NPD more than 25 years to re-enter state parliaments in Saxony (2004) and Mecklenburg-Western Pomerania (2006). In both states, the party got reelected in the subsequent elections, in 2009 and 2011, respectively. In 2001, the federal parliament brought in a claim to the German Constitutional Court to forbid the NPD due to its anti-constitutional program. The claim was turned down in 2003 because the NPD's leadership was infiltrated by domestic intelligence services agents, which caused legal problems. A second claim to forbid the party, filed on December 7th 2015, was denied by the constitutional judges on January 17th 2017.

The DVU was founded by publisher Gerhard Frey as an informal association in 1971. Frey published far-right newspapers such as the German National Newspaper (DNZ) and a number of books with the goal of mitigating Germany's role in WWII. His reputation as a publisher of far-right material helped Frey to become an influential player in the German postwar extreme-right scene (Mudde (2000)). In 1986, Frey took it one step further starting his own far-right party German List (*Deutsche Liste*). After some name changes, the party became known as German People's Union (DVU) from 1987 on. Since its foundation, the DVU got parliamentary seats in the state assemblies of Brandenburg (1999, 2004), Bremen (1991, 1999, 2003, 2007), Schleswig-Holstein (1992), and Saxony-Anhalt (1998). In 2010, the DVU merged with the NPD.

The Republicans (Die Republikaner) were founded in 1983 as an ultraconservative breakaway from the Christian Democratic Union (CDU) and the Christian Social Union of Bavaria (CSU). Under their leader, Franz Schönhuber (who also ran as a candidate for the DVU and NPD in his later political career), the party moved further to the extreme right by propagating a xenophobic view on immigrants, and particularly asylum seekers. Compared to the NPD and DVU, the Republicans were considered to be less openly extreme right which helped it secure votes from the ultraconservative clientele. The REP got parliamentary seats in Berlin's senate (1989) and the state parliament of Baden-Wuerttemberg (1992, 1996).

Online Appendix K Descriptive Statistics

Table [Online Appendix Table 3](#) provides descriptive statistics for our main variables. The table is organized in the following way: Each row presents the distribution of one variable, sliced into its 25th percentile, median, and 75th percentile. Columns 1–3 do this for Period 1 from 1987–1998, and columns 4–6 for Period 2 from 1998–2009. T_{it} is defined in units of 1,000€ per worker in constant 2005 prices.

Table Online Appendix Table 3: The Core Variables in 1987–1998 and in 1998–2009

		(1)	(2)	(3)	(4)	(5)	(6)
		Period 1 (1987-1998),			Period 2 (1998-2009),		
percentile:		25th	median	75th	25th	median	75th
T_{it}		-0.264	0.068	0.521	-1.222	-0.663	-0.144
$\widehat{T}_{it}$	(instrumented with Z_{it})	-0.068	0.143	0.402	-1.150	-0.574	-0.113
Y_{it} :	Δ Turnout	-0.034	-0.020	-0.012	-0.167	-0.128	-0.095
	Δ Vote Share CDU/CSU	-9.234	-7.659	-5.730	-4.493	-2.258	0.620
	Δ Vote Share SPD	4.120	6.472	8.248	-19.904	-17.936	-16.079
	Δ Vote Share FDP	-2.933	-2.188	-1.467	6.942	8.459	9.820
	Δ Vote Share Green	-1.779	-1.282	-0.616	2.513	3.673	4.770
	Δ Share Extreme-Right	1.520	2.086	3.099	-1.525	-1.021	-0.478
	Δ Share Far-Left	0.677	0.908	1.165	5.688	7.078	8.373
	Δ Share Small Parties	1.211	1.487	1.796	0.716	1.514	2.525
M_{it} :	$\Delta \log(\text{Total Employment})$	-0.067	0.001	0.081	-0.110	-0.044	0.021

Notes: Period one covers 1987–1998, period two covers 1998–2009. The table displays the 25th percentile, median, and 75th percentile of T_{it} , the voting outcomes Y_{it} , and manufacturing's share of employment M_{it} .

Online Appendix L Robustness and Further Results

Online Appendix L.1 Labor Market Outcomes

Online Appendix L table Online Appendix Table 4 presents the OLS results corresponding to the paper's table 3.

Table Online Appendix Table 4: OLS Version of Table 3

	(1) Baseline OLS	(2) + Industry OLS	(3) +Socio OLS	(4) + Voting OLS
$\Delta \log(\text{Total Employment})$	-0.013*** (-3.138)	-0.011** (-2.514)	-0.009** (-2.070)	-0.009* (-1.919)
$\Delta \text{Share Manufacturing Employment}$	-0.502*** (-3.348)	-0.524*** (-3.486)	-0.496*** (-3.289)	-0.502*** (-3.362)
$\Delta \log(\text{Mean Manufacturing Wage})$	-0.003** (-2.122)	-0.004** (-2.262)	-0.003** (-2.094)	-0.003** (-2.152)
$\Delta \log(\text{Mean Non-Manufacturing Wage})$	-0.001 (-0.934)	-0.001 (-0.853)	-0.000 (-0.351)	-0.000 (-0.433)
$\Delta \text{Share Unemployment}$	0.089* (1.659)	0.095 (1.617)	0.102* (1.732)	0.125*** (2.674)
Period-by-region FE	Yes	Yes	Yes	Yes
Observations	730	730	730	730

Notes: T-statistics reported, standard errors are clustered at the level of 96 commuting zones, *** p<0.01, ** p<0.05, * p<0.1.

Online Appendix L.2 Additional Results on the Vote Share Table 4

Online Appendix L table Online Appendix Table 5 presents the OLS results corresponding to the paper's table 4.

Table Online Appendix Table 5: OLS Version of Table 4

	(1) Baseline OLS	(2) + Industry OLS	(3) +Socio OLS	(4) + Voting OLS	(5) Standard. OLS
Δ Turnout	0.004*** (2.932)	0.004*** (3.059)	0.003** (2.337)	0.003** (2.430)	0.040** (2.430)
<i><u>Established Parties:</u></i>					
Δ Vote Share CDU/CSU	-0.081 (-1.015)	-0.113 (-1.423)	-0.062 (-0.963)	-0.067 (-1.020)	-0.016 (-1.020)
Δ Vote Share SPD	-0.037 (-0.416)	-0.044 (-0.471)	0.061 (0.884)	0.062 (0.929)	0.005 (0.929)
Δ Vote Share FDP	0.094** (1.971)	0.105** (2.398)	0.081* (1.805)	0.088** (2.097)	0.016** (2.097)
Δ Vote Share Green Party	0.046 (1.221)	0.063* (1.755)	0.062* (1.835)	0.068** (2.042)	0.024** (2.042)
<i><u>Non-established Parties</u></i>					
Δ Vote Share Extreme-Right Parties	0.038* (1.703)	0.036 (1.522)	-0.009 (-0.483)	-0.004 (-0.240)	-0.002 (-0.240)
Δ Vote Share Far-Left Parties	-0.108* (-1.669)	-0.109 (-1.597)	-0.138** (-2.159)	-0.153** (-2.491)	-0.039** (-2.491)
Δ Vote Share Other Small Parties	0.048 (1.586)	0.062** (2.186)	0.003 (0.138)	0.007 (0.259)	0.005 (0.259)
Period-by-region FE	Yes	Yes	Yes	Yes	Yes
Observations	730	730	730	730	730

Notes: T-statistics reported, standard errors are clustered at the level of 96 commuting zones, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Online Appendix M Constructing Gravity Residuals

Gravity-residuals can be obtained from the residuals of the regression

$$\log(EX_{djt}^{CE-O}) - \log(EX_{djt}^{G-O}) = \alpha_d + \alpha_j + \epsilon_{djt}^{IM}, \quad (149)$$

where $\log(EX_{djt}^{CE-O})$ are industry j 's log export values from China and Eastern Europe to destination market d , $\log(EX_{djt}^{G-O})$ are German industries' exports to the same countries, α_d are destination-market and α_j are industry-fixed effects.⁷⁸ ϵ_{djt}^{IM} thus captures CE 's competitive advantage over Germany at time t in destination market d and industry j . Averaging residuals ϵ_{djt}^{IM} over destination markets d and taking first differences provides a measure for overall changes in CE 's comparative advantage over time. Exponentiating this term and multiplying it with Germany's start-of-period imports from CE gives rise to $\Delta IM_{Gjt}^{grav} = IM_{Gjt-1} \times \exp^{\bar{\epsilon}_{jt}^{IM} - \bar{\epsilon}_{jt-1}^{IM}}$, which is a counterfactual measure of changes in German industries' import exposure that is solely driven by CE 's increasing comparative advantage.

Conversely, ΔEX_{Gjt} increases due to better access to the CE markets and to German-specific supply conditions. While German-specific supply conditions will affect German exports in general, the relative attractiveness of CE markets over other export destinations should be independent of German-specific effects. Thus, changes in German industries' exports to China and Eastern Europe in relation to German industries' exports to other countries O provides an exogenous measure ΔEX_{Git}^{grav} for ΔEX_{Gjt} . It can be obtained from the residuals of the regression

$$\log(EX_{djt}^{G-CE}) - \log(EX_{djt}^{G-O}) = \alpha_d + \alpha_j + \epsilon_{djt}^{EX}, \quad (150)$$

where $\log(EX_{djt}^{G-CE})$ are industry j 's log export values from Germany to China and Eastern Europe, $\log(EX_{djt}^{G-O})$ are German industries' exports to other countries, and α_d and α_j are again destination-country and industry-fixed effects. ϵ_{djt}^{EX} now captures CE 's relative attractiveness over other sales markets at time t in destination market d and industry j . Averaging residuals ϵ_{djt}^{EX} over destination markets d and taking first differences provides a measure for overall changes in the attractiveness of Chinese and Eastern European sales markets over time. Exponentiating this term and multiplying it with Germany's start-of-period exports to CE gives rise to $\Delta EX_{Gjt}^{grav} = EX_{Gjt-1} \times \exp^{\bar{\epsilon}_{jt}^{EX} - \bar{\epsilon}_{jt-1}^{EX}}$, which is a counterfactual measure of changes in German industries' export exposure that is solely driven by CE 's increasing attractiveness as sales market.

References

- Acemoglu, Daron and Pascual Restrepo, "Robots and Jobs: Evidence from US Labor Markets.," *MIT Unpublished Mimeo.*, 2017.
- Art, David, "Reacting to the Radical Right Lessons from Germany and Austria," *Party Politics*, 2007, 13 (3), 331–349.
- Arzheimer, K., "Contextual Factors and the Extreme Right Vote in Western Europe, 1980-2002," *American Journal of Political Science*, 2009, 53 (2), 259–275.

⁷⁸Since many CE countries did not report trade data in the late 1980s and early 1990s, we use imports from CE and Germany reported by other countries O to measure Germany's and CE 's exports to O .

Table Online Appendix Table 6: Individual-Level Analysis

	(1)	(2)	(3)	(4)
	1990-1998	1998-2009	Low-Skill & Manuf., 1998-2009	Low-Skill & Not Manuf., 1998-2009
<u><i>Established Parties:</i></u>				
Would Vote CDU/CSU	-0.025 (-0.743)	0.002 (0.227)	-0.006 (-0.350)	0.003 (0.257)
Would Vote SPD	0.027 (0.761)	-0.019** (-2.217)	0.001 (0.031)	-0.022** (-2.352)
Would Vote FDP	-0.038 (-0.720)	0.015 (1.177)	0.002 (0.116)	0.021 (1.431)
Would Vote Green Party	0.019 (0.409)	0.016 (1.295)	0.007 (0.363)	0.007 (0.565)
<u><i>Non-Established Parties:</i></u>				
Would Vote Extreme-Right Parties	0.029 (0.735)	0.028* (1.802)	0.088** (2.013)	0.016 (1.035)
Would Vote Far-Left Parties	-0.008 (-0.670)	-0.005 (-0.751)	0.026 (1.579)	-0.010 (-1.043)
Would Vote Other Small Parties	0.018 (0.340)	-0.012 (-1.072)	-0.048* (-1.674)	-0.001 (-0.042)
Period-by-region F.E.	Yes	Yes	Yes	Yes
Observations	3,694	5,975	1,168	3,817

Notes: (a) Columns 1–2 split the the paper’s table A2 by period (3,694 + 5,975= 9,669). The results are driven entirely by period 2, i.e. after Germany’s labor markets were de-regulated. No part of the political spectrum responds in period 1. In period 2, SPD support is reduced in response to import exposure and support for the extreme right goes up. In columns 3–4, we focus on the second period, which sharpens the results of the worker-type breakdowns in the paper’s table A2. (b) Standard errors are clustered at the region level. *** p<0.01, ** p<0.05, * p<0.1.

- Autor, David and David Dorn**, “The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market,” *American Economic Review*, 2013, 103 (5), 1553–1597.
- , —, and **Gordon H Hanson**, “Untangling trade and technology: Evidence from local labour markets,” *The Economic Journal*, 2015, 125 (584), 621–646.
- Bender, Stefan, Anette Haas, and Christoph Klose**, “IAB Employment Subsample 1975-1995 Opportunities for Analysis Provided by the Anonymised Subsample,” *IZA Discussion Paper 117*, 2000.
- Bloom, Nicholas, Mirko Draca, and John Van Reenen**, “Trade induced technical change? The impact of Chinese imports on innovation, IT and productivity,” *The Review of Economic Studies*, 2016, 83 (1), 87–117.
- Burgess, S., R. M. Daniel, A. S. Butterworth, and S. G. Thompson**, “Network Mendelian randomization: using genetic variants as instrumental variables to investigate mediation in causal pathways,” *International Journal of Epidemiology*, 2015, 44 (2), 484–495.
- Che, Yi, Yi Lu, Justin R Pierce, Peter K Schott, and Zhigang Tao**, “Does Trade Liberalization with China Influence US Elections?,” Technical Report, National Bureau of Economic Research 2016.
- Dauth, Wolfgang, Sebastian Findeisen, and Jens Suedekum**, “The Rise of the East and the Far East: German Labor Markets and Trade Integration,” *Journal of European Economic Association*, 2014, 12 (6), 1643–1675.
- Dunn, G. and R. Bentall**, “Modelling treatment-effect heterogeneity in randomized controlled trials of complex interventions (psychological treatments),” *Statistics in Medicine*, 2007, 26 (26), 4719–4745.
- Falck, Oliver, Robert Gold, and Stephan Heblich**, “E-lections: Voting Behavior and the Internet,” *American Economic Review*, 2014, 104 (7), 2238–65.
- , **Stephan Heblich**, and **Anne Otto**, “Agglomerationsvorteile in der Wissensgesellschaft: Empirische Evidenz für deutsche Gemeinden,” *ifo Schnelldienst*, 2013, 66 (3), 17–21.
- Falk, Armin, Andreas Kuhn, and Josef Zweimüller**, “Unemployment and Right-wing Extremist Crime,” *The Scandinavian Journal of Economics*, 2011, 113 (2), 260–285.
- Frank, Thomas**, “Millions of Ordinary Americans Support Donald Trump. Here’s Why,” *The Guardian*, March 7th 2016.
- Frey, Carl B. and Michael A. Osborne**, “The Future of Employment: How Susceptible are Jobs to Computerisation?,” *Oxford Martin School Unpublished Mimeo.*, 2017.
- Frölich, Markus and Martin Huber**, “Direct and indirect treatment effects: causal chains and mediation analysis with instrumental variables,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2017, pp. n/a–n/a.
- Geneletti, S.**, “Identifying direct and indirect effects in a non-counterfactual framework,” *Journal of the Royal Statistical Society B*, 2007, 69 (2), 199–215.
- Gennetian, L. A., J. Bos, and P. Morris**, “Using Instrumental Variables Analysis to Learn More from Social Policy Experiments,” MDRC Working Papers on Research Methodology, MDRC (Manpower Demonstration Research Corporation) 2002.

- Grumke, Thomas**, *The Extreme Right in Europe*, Vandenhoeck & Ruprecht, 2012.
- Have, T. R. Ten, M. M. Joffe, K. G. Lynch, G. K. Brown, S. A. Maisto, and A. T. Beck**, “Causal mediation analyses with rank preserving models,” *Biometrics*, September 2007, 63 (3), 926–934.
- Heckman, James J.**, “The Principles Underlying Evaluation Estimators with an Application to Matching,” *Annales d’Economie et de Statistiques*, 2008, 91–92, 9–73.
- Heckman, James J and Rodrigo Pinto**, “Econometric mediation analyses: Identifying the sources of treatment effects from experimentally estimated production technologies with unmeasured and mismeasured inputs,” *Econometric reviews*, 2015, 34 (1-2), 6–31.
- Imai, Kosuke, Luke Keele, and Te Yamamoto**, “Identification, Inference and Sensitivity Analysis for Causal Mediation Effects,” *Statistical Science*, 2010, 25 (1), 51–71.
- , —, —, **Dustin Tingley, and Teppei Yamamoto**, “Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies,” *American Political Science Review*, 2011, 105, 765–789.
- , —, —, —, and —, “Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies,” *American Political Science Review*, 2011, 105 (4), 765–789.
- Jhun, M. A.**, “Epidemiologic approaches to understanding mechanisms of cardiovascular diseases: genes, environment, and DNA methylation.” PhD dissertation, University of Michigan, Ann Arbor 2015.
- Jun, Sung Jae, Joris Pinkse, Haiqing Xu, and Neşe Yildiz**, “Multiple Discrete Endogenous Variables in Weakly-Separable Triangular Models,” *Econometrics*, 2016, 4 (1).
- Krueger, Alan B and Jörn-Steffen Pischke**, “A Statistical Analysis of Crime Against Foreigners in Unified Germany,” *Journal of Human Resources*, 1997, 32 (1), 182–209.
- Lubbers, Marcel and Peer Scheepers**, *European Sociological Review*, 2001, 17 (4), 431–449.
- Mocan, Naci H and Christian Raschke**, “Economic Well-being and Anti-Semitic, Xenophobic, and Racist Attitudes in Germany,” *National Bureau of Economic Research Working Paper* 20059, 2014.
- Mudde, Cas**, *The Ideology of the Extreme Right*, Manchester University Press, 2000.
- Mughan, Anthony, Clive Bean, and Ian McAllister**, “Economic globalization, job insecurity and the populist reaction,” *Electoral Studies*, 2003, 22 (4), 617–633.
- Pearl, Judea**, “The Mediation Formula: A Guide to the Assessment of Causal Pathways in Non-linear Models,” 2011. Forthcoming in *Causality: Statistical Perspectives and Applications*.
- Petersen, M. L., S. E. Sinisi, and M. J. Van der Laan**, “Estimation of direct causal effects,” *Epidemiology*, 2006, 17, 276–284.
- Powdthavee, Nattavudh**, “Does Education Reduce Blood Pressure? Estimating the Biomarker Effect of Compulsory Schooling in England,” Discussion Paper 09/14, University of York, Department of Economics, York, UK 2009.

- Robins, J. M.**, "Semantics of causal DAG models and the identification of direct and indirect effects.," in N. L. P. J. Green, Hjort and S. Richardson, eds., *Highly Structured Stochastic Systems*, MR2082403, Oxford: Oxford University Press, 2003, pp. 70–81.
- Robins, James M. and Sander Greenland**, "Identifiability and Exchangeability for Direct and Indirect Effects," *Epidemiology*, 1992, 3 (2), 143–155.
- Rosenbaum, Paul R. and Donald B. Rubin**, "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, April 1983, 70 (1), 41–55.
- Rubin, D. B.**, "Direct and indirect causal effects via potential outcomes (with discussion)," *Scandinavian Journal of Statistics*, 2004, 31, 161–170.
- Small, D. S.**, "Mediation analysis without sequential ignorability: using baseline covariates interacted with random assignment as instrumental variables," *Journal of Statistical Research*, 2012, 46 (2), 91–103.
- Sommer, Bernd**, "Anti-capitalism in the name of ethno-nationalism: ideological shifts on the German extreme right," *Patterns of Prejudice*, 2008, 42 (3), 305–316.
- Stöss, Richard**, "Rechtsextremismus im Wandel," Technical Report, Friedrich Ebert Stiftung 2010.
- The Economist**, "The New Political Divide," July 30th, 2016.
- Voigtländer, Nico and Hans-Joachim Voth**, "Taught to Hate: Nazi Indoctrination and Anti-Semitic Beliefs in Germany," *Proceedings of the National Academy of Sciences*, 2015, *Forthcoming*.
- Yamamoto, T.**, "Identification and estimation of causal mediation effects with treatment noncompliance," March 2014. Manuscript. Department of Political Science, Massachusetts Institute of Technology, Cambridge.